KU LEUVEN

ARENBERG DOCTORAL SCHOOL
FACULTY OF SCIENCE

Multi-scale aquifer characterization

from outcrop analogue, direct-push and borehole investigations towards improved groundwater flow models

Bart Rogiers

Supervisory Committee:
Prof. Dr. Okke Batelaan
Prof. Dr. Ir. Marijke Huysmans
Prof. Dr. Ir. Alain Dassargues
Dr. Ir. Dirk Mallants
Dr. Matej Gedeon

Dissertation presented in partial fulfilment of the requirements for the degree of Doctor of Science, Geology

November 2013

MULTI-SCALE AQUIFER CHARACTERIZATION

FROM OUTCROP ANALOGUE, DIRECT-PUSH AND BOREHOLE INVESTIGATIONS TOWARDS IMPROVED GROUNDWATER FLOW MODELS

Bart ROGIERS

Supervisors:
Prof. Dr. O. Batelaan
Prof. Dr. Ir. M. Huysmans
Prof. Dr. Ir. A. Dassargues
Université de Liège

Dissertation presented in partial fulfilment of the requirements for the degree of Doctor of Science, Geology

Members of the Examination Committee:
Prof. Dr. N. Vandenberghe, president
Prof. Dr. R. Swennen
Prof. Dr. Ir. J. Vanderborght
Dr. Ir. D. Mallants
CSIRO Land and Water
Dr. M. Gedeon
SCK•CEN
Prof. Dr. P. Dietrich
Helmholtz Centre for Environmental Research - UFZ

November 2013

© 2013 KU Leuven, Science, Engineering & Technology
v.u. Leen Cuypers, Arenberg Doctoraatsschool, W. de Croylaan 6, 3001 Heverlee

Alle rechten voorbehouden. Niets uit deze uitgave mag worden vermenigvuldigd en/of openbaar gemaakt worden door middel van druk, fotokopie, microfilm, elektronisch of op welke andere wijze ook zonder voorafgaandelijke schriftelijke toestemming van de uitgever.

All rights reserved. No part of the publication may be reproduced in any form by print, photoprint, microfilm, electronic or any other means without written permission from the publisher.

ISBN 978-90-8649-672-3
D/2013/10.705/87
ISSN 0250-7803
Aardkundige Mededelingen 42

Acknowledgements

These really have been four amazing years. In october 2009 I started exploring the Campine area with my first car, and soon I was travelling abroad and overseas to learn from renowned scientists and participate at international conferences. The outcome of this period is of course not my own achievement, but the work of many hands.

In the first place, I would like to thank the Belgian Nuclear Research Centre, SCK•CEN, and its scientific committee for giving me the opportunity to perform this PhD research, and giving me plenty of opportunities to present it to and interact with the international community.

Secondly, I wish to thank my promoter, Prof. Dr. Okke Batelaan, and the other members of my supervisory committee, Dr. Dirk Mallants, Dr. Matej Gedeon, Prof. Dr. Marijke Huysmans and Prof. Dr. Alain Dassargues, for their continuing support, enlighting discussions and critical remarks during the last four years. I certainly hope we can continue our interactions in the years to come.

Many thanks also go to the other members of the Examination Committee, Prof. Dr. Noël Vandenberghe, Prof. Dr. Jan Vanderborght, Prof. Dr. Rudy Swennen and Prof. Dr. Peter Dietrich, for their valuable comments which greatly improved the quality of the manuscript.

For their contributions to the scientific output of this work I like to additionally acknowledge my other co-authors Dr. Koen Beerten, Dr. Thomas Vienken, Dr. Eric Laloy, Pieter Winters, Tuur Smekens and Marco Schiltz for their input.

I also like to thank Luk Peeters for useful discussions concerning the groundwater flow model refinement and his valuable comments on different chapters of this manuscript. Valuable comments were also received from Dr. Elise Bekele, Dr. Wendy Welsh, Andrew Taylor, Dr. Brian Smerdon and Prof. Dr. Kristine Walraevens, who reviewed certain chapters of this manuscript.

Though many data have been gathered in the framework of this PhD, with the people mentioned above, this research makes use of many other datasets for which a number of people and organizations have to be thanked. I am very grateful to ONDRAF/NIRAS[1], the Belgian Agency for Radioactive Waste and Enriched Fissile Materials, for providing the CPT and borehole data, granting access to the borehole cores and supporting the hydraulic direct push campaign. Laurent Wouters is thanked especially, for his continuing interest in and support for this work.

The colleagues from UFZ Leipzig, Dr. Thomas Vienken, Markus Tinter, Manuel Kreck and Helko Kotas, are thanked for performing the hydraulic direct push campaign, and interpretation of the raw data. Uwe Schneidewind also deserves a special mention here. His knowledge and enthusiasm on direct push technologies has definitely contributed in making this campaign a succes.

Furthermore, I like to thank Dr. Katrijn Vandersteen, Rieko Adriaens and Koen Vos for their assistance during the field work, and Serge Labat and Frans Slegers for their help with handling the borehole cores.

Sibelco and Terca Nova are acknowledged for the permission to access their quarries, and Via Kempen for the permission to access the Kempense Noord-Zuidverbinding road construction yard.

Working both at SCK•CEN and KU Leuven, I have been blessed with a large number of colleagues who I would like to thank for their enjoyable company during the lunch breaks/talks, the seminars at the Mariagaarde castle, and conferences we attended together. Benedicta Ronchi and Mathias Possemiers deserve a special mention for relieving me

[1] Findings and conclusions in this manuscript are those of the authors and do not necessarily represent the official position of ONDRAF/NIRAS.

several times of the practical sessions or excursion of the KU Leuven hydrogeology course.

Finally, I would like to thank my family and friends for their support. My deepest gratitude goes to my parents, Freddy and Marina, who have given me the chance to study geology and further supported me during the last four years, and Nele, my fiancée, who has always encouraged me, and definitely helped me bear the workload during these last few hectic months.

Abstract

Groundwater flow and solute transport modelling are affected by different kinds of uncertainty including spatial variability in aquifer properties such as hydraulic conductivity (*K)*. This spatial variability is often present at different scales, and consequently, effective *K* values are very much scale-dependent. While regional-scale models often use effective *K* values obtained by inverse modelling, the small-scale variability has to be accounted for as well, as it has been shown that even submeter-scale heterogeneity can have important consequences on solute transport in aquifers. To support decision making related to environmental impact assessment for waste disposal sites or sites contaminated by point sources, appropriate subsurface characterization and modelling tools are thus required for accounting for subsurface heterogeneity, possibly observed at multiple spatial scales. These tools can further be used to quantify the uncertainty associated with groundwater flow and solute transport, to underpin strategies for long-term groundwater quality monitoring at disposal or contaminated sites, or for developing groundwater remediation schemes.

In this work, a methodology is developed for efficient multi-scale subsurface characterization and integration of the gathered data in a stochastic regional groundwater flow and solute transport modelling approach. Different kinds of aquifer characterization technologies have to be combined to cover the centimetre- to the kilometre-scale, and to make optimal use of common or easily gathered secondary data. Additionally, for accounting for secondary data, the development of a set of tools for data calibration and interpretation is required. This is achieved by using different kinds of measurements from outcrop analogues, borehole and direct-push investigations, and by using innovative methods and techniques to obtain a sound framework for integrating all data.

The case study that we use throughout the thesis is an area of ~60 km² in Mol/Dessel, Belgium, of which the subsurface consists of a succession of dipping lithostratigraphical units with varying degrees of heterogeneity, all part of the Neogene aquifer. We make use of different previous hydrogeological studies that were performed in the framework of the ONDRAF/NIRAS radioactive waste disposal programmes, and perform additional site characterization.

The assessment of outcrops as analogues for the subsurface sediments shows that very useful quantitative and qualitative information can be retrieved from outcrops, but a systematic bias seems to exist between surface and subsurface data (lower *K* for the latter). The relative differences however seem to be valid, as is the amount of spatial heterogeneity. For including small-scale heterogeneity in the developed large-scale transport model, we make use of dispersivities estimated from the outcrop characterization.

The three types of secondary *K* data used in this work are air permeability, grain size and cone penetration test data. For each of these we use a data-driven modelling approach to obtain estimates of *K*. The complexity of these approaches ranges from a simple linear model to the combination of artificial neural networks with general likelihood estimation. Each time, the site-specific estimates prove to be superior to existing models from literature.

The conditioning of a regional groundwater flow model on borehole and geotechnical and hydraulic direct push data improves the model performance considerably. We use a data-driven approach for the hydrostratigraphy, by invoking a non-stationary multivariate geostatistical framework for conditioning the model. A combination of different McMC algorithms is used to estimate the uncertainty of the flow field, for the quantification of uncertainty on the corresponding solute transport. The combined algorithm is tested both in McMC sampling and optimization mode. The latter is clearly more efficient for CPU-intensive models, but only provides an approximation of the posterior.

Solute transport simulations in the framework of surface disposal of radioactive waste, based on the obtained flow solutions, suggest that the reference model that considers homogeneous lithostratigraphical units produces conservative results in terms of maximum concentrations within the solute plume. For more detailed results, transport simulations using another advection solution scheme and finer numerical discretization are however recommended.

Samenvatting

Het modelleren van grondwaterstroming en het transport van opgeloste stoffen wordt beïnvloed door verschillende soorten van onzekerheid, waaronder de ruimtelijke variabiliteit van aquifer parameters zoals de hydraulische conductiviteit (*K*). Zulke ruimtelijke variabiliteit is vaak aanwezig op verschillende schalen, en bij gevolg zijn effectieve *K* waarden zeer schaal-afhankelijk. Hoewel regionale modellen vaak gebruik maken van effectieve *K* waarden door inverse modellering, moet er ook rekening gehouden worden met de kleinschalige variabiliteit, vermits zelfs sub-meter schaal heterogeniteit belangrijke effecten kan hebben op het transport van opgeloste stoffen. Ter ondersteuning van de besluitvorming in verband met de rapportage van effecten op het milieu voor bergingssites of locaties verontreinigd door puntbronnen zijn er verschillende tools nodig voor de karakterisatie alsook de modellering van de ondergrond zodat de heterogeniteit in de ondergrond, die mogelijk op verschillende schalen aanwezig is, in rekening gebracht kan worden. Deze tools kunnen dan verder gebruikt worden om de onzekerheid in verband met grondwaterstroming en het transport van opgeloste stoffen te kwantificeren, om strategieën voor lange-termijn monitoring van de grondwaterkwaliteit op bergings- of verontreinigde sites, of voor het saneren van grondwater te onderbouwen.

In dit werk wordt een methodologie ontwikkeld voor efficiënte multischaal karakterisatie van de ondergrond, alsook de integratie van de bekomen gegevens in stochastische benadering van het modelleren van grondwaterstroming en het transport van opgeloste stoffen. Verschillende technologiën voor het karakteriseren van aquifers dienen gecombineerd te worden om de centimeter- tot de kilometer-schaal te dekken, alsook gebruik te kunnen maken van frequent of gemakkelijk verzamelde secundaire gegevens. Bovendien is de ontwikkeling van een set van tools voor data-calibratie en interpretatie vereist om zulke secundaire gegevens te kunnen verwerken. Dit wordt gerealiseerd door het gebruik van verschillende soorten metingen voor analoge ontsluitingen, boringen, en direct-push testen, en door innovatieve methodes en technieken te combineren voor de integratie van alle gegevens.

De case study die we gebruiken in dit werk is een gebied van ~60 km² in Mol/Dessel, België, waarvan de ondergrond bestaat uit een opeenvolging van gekantelde lithostratigrafische eenheden met verschillende mate van heterogeniteit, allemaal onderdeel van de Neogene aquifer. We maken gebruik van verschillende hydrogeologische studies die werden uitgevoerd in het kader van de projecten van ONDRAF/NIRAS voor de berging van radioactief afval, en doen daarnaast ook additionele karakterisatie van het studiegebied.

Het gebruik van ontsluitingen als analogen voor de sedimenten in de ondergrond bleek zeer nuttige kwalitatieve alsook kwantitatieve informatie op te leveren. Een systematische afwijking blijkt wel te bestaan tussen de ontsluitingen en de gegevens representatief voor de ondergrond (lagere *K* waarden voor de laatste). De relatieve verschillen tussen ontsluitingen blijken wel correct, net als de mate van ruimtelijke variabiliteit. Om het effect van de kleinschalige heterogeniteit in het grootschalige transport model te brengen maken we gebruik van dispersiviteiten, geschat op basis van de karakterisatie van ontsluitingen.

De drie types secundaire *K* data die gebruikt werden in dit werk zijn lucht permeabiliteit, korrelgrootte verdelingen en sonderingen. Voor elk van deze datatypes gebruiken we data-gedreven modellering om *K* te schatten. De complexiteit van de verschillende benaderingen varieert van het gebruik van een eenvoudig lineair model tot de combinatie van artificiële neurale netwerken met de general likelihood onzekerheidsschatting-methodologie. De site-specifieke schattingen van *K* bleken steeds

betere resultaten op te leveren dan bestaande modellen uit de literatuur.

De conditionering van een regionaal grondwaterstromingsmodel op gegevens van boringen, sonderingen en hydraulisch direct push testen verbetert de performantie van dat model aanzienlijk. We gebruiken eveneens een data-gedreven aanpak voor de hydrostratigrafie, door het gebruik van niet-stationaire multivariate geostatistiek voor het conditioneren van het model. Een combinatie van verschillende McMC algorithmes werd gebruikt voor het schatten van de onzekerheid op de grondwaterstroming, en de overeenkomstige simulaties van het transport van opgeloste stoffen. Het gecombineerde algorithme werd zowel in McMC als in optimalisatie mode getest. Deze laatste is duidelijk efficiënter voor CPU-intensieve modellen, maar geeft slechts een benadering van de werkelijke onzekerheid.

De simulatie van het transport van opgeloste stoffen in het kader van de oppervlaktebeging van radioactief afval, op basis van de grondwaterstroming uit het stromingsmodel, suggereert dat het referentiemodel met homogene lithostratigrafische eenheden conservatieve resultaten oplevert in termen van maximale concentraties binnen de verontreinigingspluim. Voor meer gedetailleerde resultaten worden simulaties met een ander schema voor het oplossen van de advectie term, alsook het verfijnen van de numerische discretisatie aangeraden.

Contents

Chapter 1

General introduction

1.1. The research question

Groundwater flow and solute transport modelling are affected by different kinds of uncertainty (Nilsson *et al.* 2007; Rojas *et al.* 2010; Refsgaard *et al.* 2011), including spatial variability in aquifer properties such as hydraulic conductivity (*K*) (*e.g.* Koltermann & Gorelick 1996; De Marsily *et al.* 2005; Sudicky *et al.* 2010; Blouin *et al.* 2012). This spatial variability is often present at different scales, and consequently, effective *K* values are very much scale-dependent (Rovey & Cherkauer 1995; Boschan & Nœtinger 2012). While regional-scale models often use effective *K* values obtained by inverse modelling, the small-scale variability has to be accounted for as well, as it has been shown that even submeter-scale heterogeneity can have important consequences on solute transport in aquifers (*e.g.* Huysmans & Dassargues 2009; Ronayne *et al.* 2010).

To support decision making related to environmental impact assessment for waste disposal sites or sites contaminated by point sources, appropriate subsurface characterization and modelling tools are required for accounting for subsurface heterogeneity, possibly observed at multiple spatial scales (Swartjes 2011; Bailey *et al.* 2011; de Barros *et al.* 2011). These tools can further be used to quantify the uncertainty associated with groundwater flow and solute transport (*e.g.* Mariethoz *et al.* 2010; Hill *et al.* 2012; Laloy *et al.* 2013; Meyer *et al.* 2013), to underpin strategies for long-term groundwater quality monitoring at disposal or contaminated sites (*e.g.* Reed & Minsker 2004; Li & Chan Hilton 2007) or for developing groundwater remediation schemes (*e.g.* Quinnan *et al.* 2010; Yang *et al.* 2013).

Characterization of aquifer heterogeneity is a challenging task. Given the existence of multi-scale subsurface heterogeneity, different characterization techniques are used at different spatial scales. The use of pumping tests and lab analyses of core samples leads to very different measurement support volumes, which complicates the integration of different data types considerably (*e.g.* Tartakovsky *et al.* 2000; Zlotnik *et al.* 2000). To complement direct *K* measurements, additional soft data like grain size or geophysical/geotechnical properties of the subsurface are frequently available. They can complement the hard *K* data if reliable methods are available to relate primary and secondary variables (*e.g.* Huysmans & Dassargues 2005; Vienken & Dietrich 2011; Jiang *et al.* 2013). Characterization of submeter-scale heterogeneity is also not straightforward, certainly not in the horizontal direction.

The systematic use of small-scale measurement supports seems therefore appealing to enable straigthforward integration of different data types. Detailed characterization of borehole cores might be used for this purpose, with the classical expensive and time-consuming drilling approach. Direct push technologies (Dietrich & Leven 2006) represent another, more efficient option for high-resolution characterization of the shallow subsurface with small-scale measurement support. Geotechnical direct push methods, and more specific cone penetration tests (CPTs) (Lunne *et al.* 1997) have been around for ~60 years, and allow nowadays for profiling the shallow subsurface at a speed of 2 cm/s in unconsolidated sedimentary formations.

Several studies have already investigated correlations between CPT data and hydrogeological parameters such as *K* (*e.g.* Tillmann *et al.* 2008; Anagnostopoulos *et al.* 2003; Robertson 2010; Van Der Wal *et al.* 2010). However, reported efforts on using geotechnical data for conditioning stochastic realizations of hydrogeological models are very

limited, though gathering of such information is usually much easier and cheaper than performing expensive drilling campaigns. Furthermore, more recent developments in the field of hydraulic direct push site characterization have enabled *in situ* measurement of *K*, at the resolution of CPT data (Dietrich *et al.* 2008; Köber *et al.* 2009; Liu *et al.* 2009; Lessof *et al.* 2010; Schmelzbach *et al.* 2011).

While these technologies are effective in high-resolution characterization of small-scale heterogeneity in the vertical direction, getting information on the horizontal spatial variability remains costly and time-consuming as many tests should be performed at short distances. Such investigations could therefore be complemented by outcrop analogue studies to provide more understanding of the sediment properties in 2D. The integration of detailed borehole core analyses (providing both hard and soft data), geotechnical and hydraulic direct push measurements, and outcrop analogue studies is therefore the subject of this research. The overall goal is to develop a methodology for efficient multi-scale subsurface characterization and integration of the gathered data in a stochastic regional groundwater flow and solute transport modelling approach. The case study that we use throughout the thesis is an area of ~60 km² in Mol/Dessel, Belgium, where different hydrogeological studies on the Neogene aquifer have been performed in the framework of *i.a.* the ONDRAF/NIRAS category A disposal programme.

1.2. The category A radioactive waste disposal programme

ONDRAF/NIRAS, the Belgian Agency for Radioactive Waste and Enriched Fissile materials, and its partners have been carrying out studies for the development of a surface facility for the disposal of low- and intermediate-level short lived radioactive waste (category A waste).

In this framework, the primary task of ONDRAF/NIRAS is to develop an integrated waste disposal project, with the goal to receive all the necessary licences and reach a binding agreement between the stakeholders for the realization of the near surface disposal facility and the complementary conditions.

The radiological impact and safety assessment for the geosphere component, which is part of the ONDRAF/NIRAS safety assessment methodology, is part of the research performed by SCK•CEN, the Belgian Nuclear Research Centre, in the framework of the cAt project. The geosphere is the main pathway for radionuclides from source to receptor. Processes in the geosphere additionally decrease the radionuclide concentrations that escape the surface disposal facility, designed to isolate and contain the radionuclides from the biosphere and Man. The radiological impact and safety assessment for the geosphere is performed using groundwater flow and solute transport modelling (Gedeon & Mallants 2004, 2012; Gedeon *et al.* 2011). The heterogeneity within the Neogene aquifer has previously been considered in a very simplified manner by invoking large-scale lithostratigraphical layers that were internally homogeneous as regards to their hydrogeological parameters (Gedeon *et al.* 2011). The main scientific challenge is to quantify subsurface heterogeneity across a range of scales and determine its impact on groundwater flow and solute transport for a solute source configuration typical of a surface repository.

1.3. Aims of this work

The aim of this work is to develop a methodology for efficient multi-scale subsurface characterization and integration of the gathered data in a stochastic regional groundwater flow and solute transport modelling approach. Different kinds of aquifer characterization technologies have to be combined to cover the centimetre- to the kilometre-scale, often relying on a combination of hard (primary) and soft (secondary) data. A typical example of the combination of hard and soft data is the use of laboratory tests on borehole core samples with geophysical logs like resistivity and gamma ray (*e.g.* Huysmans & Dassargues 2005; Jiang *et al.* 2013). To make optimal use of secondary data, the development of a set of tools for data calibration and interpretation is required. We achieve this by using different kinds of measurements from outcrop analogues, borehole and direct-push investigations, and by developing innovative methods and techniques as part of a flexible framework for integrating all data. The methodology is applied to a subdomain of the Neogene aquifer including the future surface repository site for low level waste

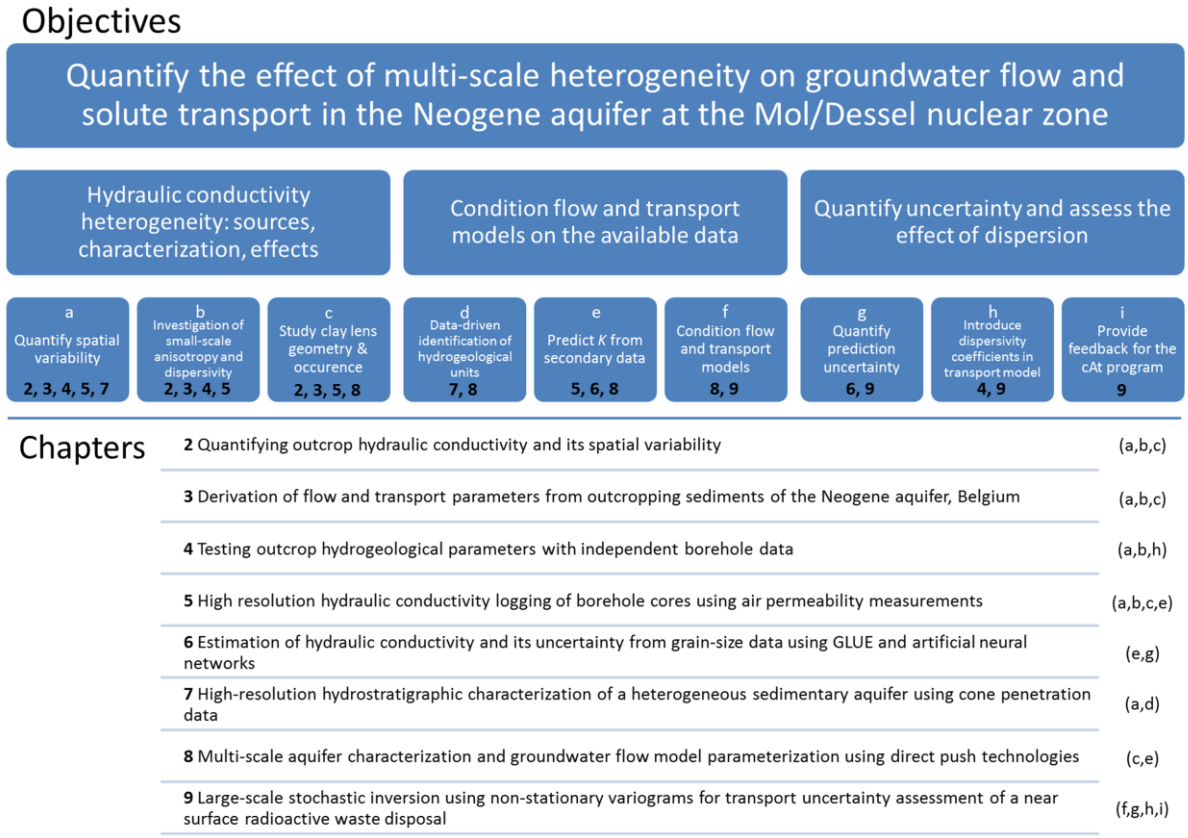

Figure 1.1. Overview of the hierarchical structure of objectives, and the corresponding chapters.

at the Mol/Dessel nuclear zone (ONDRAF/NIRAS 2010).

This study has a wide-range applicability within the groundwater management community, as it addresses key considerations about characterization of spatial variability of subsurface properties from the centimetre- to the kilometre-scale, and subsequent incorporation of hard and soft hydrogeological data in stochastic models of flow and transport.

The objectives of this PhD research are provided below and in Figure 1.1 in a hierarchical way.

Overall goal

Quantify the effect of multi-scale heterogeneity on groundwater flow and solute transport in the Neogene aquifer at the Mol/Dessel nuclear zone

The overall goal is to quantify the effect of multi-scale heterogeneity of hydraulic conductivity (*K*) on groundwater flow at the study area surrounding the Mol/Dessel nuclear site. The effect on solute transport in the framework of the cAt project has to be quantified as well. This overall goal is divided into three primay objectives, each with three secondary objectives.

Primary objective 1

Hydraulic conductivity heterogeneity: sources, characterization, effects

Detailed characterization of *K* and its heterogeneity is a prerequisite for the creation of realistic *K* field realizations. This characterization is divided into three parts: quantification of spatial variability in terms of two-point spatial covariance (or variograms), investigation of the small-scale anisotropy and dispersivity, and study of the clay lens geometry and occurrence and the corresponding upscaling.

Secondary objective a (2,3,4,5,7)[2]

Quantify spatial variability

Quantification of *K* spatial variability on the centimetre- to the kilometre-scale is required for accounting for multi-scale heterogeneity. Geostatistical variography is the main tool used for this purpose. Given the lack of clear complex sedimentary structures with important hydrogeological implications in the main

[2] Numbers between brackets provide reference to the chapters related to the objective.

lithostratigraphical units of the studied Neogene aquifer, we make use of two-point geostatistics throughout this work to capture the spatial patterns. The existence of considerable geostatistical non-stationarity throughout the aquifer, as demonstrated by the existence of different lithostratigraphical units, can be accounted for within a two-point geostatistical framework.
Multiple-point statistics are not used because of this large non-stationarity, and the lack of robust 3D facies models (training images) for the different lithostratigraphical units.

Secondary objective b (2,3,4,5)
Investigation of small-scale anisotropy and dispersivity
As *K* at any point in the subsurface is a 3D tensor, the anisotropy or ratio of horizontal to vertical *K* should be studied and the negligibility of the off-diagonal tensor components should be verified. Estimation of realistic dispersivity coefficients based on small-scale *K* variability is needed as well. Use of mechanical dispersion in a regional solute transport model allows to account for small-scale heterogeneity. Mechanical dispersion was not included in the previous solute transport simulations for the cAt project (Gedeon *et al.* 2011), as the sum of numerical and physical dispersion would definitely overestimate total dispersion, quantifying the effect of physical dispersion is required. When homogeneous *K* fields are used, macroscopic dispersion should be used for obtaining more realistic results at the scale of the solute plume. However, when heterogeneous *K* fields are used, only local-scale dispersion needs to be accounted for, as the macroscopic effect will come from the variability in the flow field. The result of both approaches should be compared.

Secondary objective c (2,3,5,8)
Study clay lens geometry and occurrence
One of the most important features within the Neogene aquifer in the study area are the thin clay lenses that mainly occur in the Kasterlee Clay aquitard (Beerten *et al.* 2010). The resolution of the existing borehole data was however not sufficient to characterize these lenses in detail. Moreover, any information on their horizontal extent was lacking. Accounting for these features is necessary to obtain reliable groundwater fluxes across the aquitard as prerequisite for determination of regional groundwater flow and solute transport features in the Neogene aquifer.

Primary objective 2
Condition flow and transport models on the available data
After characterization of *K* and its heterogeneity, the existing groundwater flow and solute transport models should be refined and extended to account for the quantified heterogeneity and honour all primary and secondary data. This conditioning is divided in three parts: the data-driven identification of hydrogeological units, prediction of *K* from secondary data, and the actual conditioning of the models.

Secondary objective d (7,8)
Data-driven identification of hydrogeological units
The development of a continuous non-stationary geostatistical framework in contrast with the current homogeneous-layer approach is required. This allows to work independent of lithostratigraphical interpretations, which are not strictly based on hydraulic properties of the sediments and are often subject to change when new data is gathered. The resulting non-stationarity should therefore be data-based and reflect the succession of the different hydraulic properties in an objective way.

Secondary objective e (5,6,8)
Predict K from secondary data
To make optimal use of available secondary data, different data-driven modelling techniques should be used, with a varying complexity depending on the amount of available data and the difficulty of the problem. We make use of air permeability measurements on outcrops and borehole cores, as well as grain size analyses.

Secondary objective f (8,9)
Condition flow and transport models
The final conditioning of the groundwater flow and solute transport models on primary and secondary data should be performed in a multi-variate non-stationary geostatistical framework, to account for all remaining uncertainties on the *K* field. Pre-processed borehole data and hydraulic direct-push investigations are used as primary data for this purpose, while geotechnical direct-push (CPT) data is used as secondary variable. An upscaling approach is invoked to move from the sample to the meter-

scale in order to make the geostatistical, groundwater flow and solute transport simulations computationally feasible.

Primary objective 3
Quantify uncertainty and assess the effect of dispersion
Once the groundwater flow and solute transport models are refined, extended, and conditioned on the available data, inverse conditioning of the models on the available observations should be performed to obtain an ensemble of realizations for uncertainty quantification, as well as a maximum likelihood estimate to investigate the magnitude of the safety margin. This objective is divided in three secondary objectives: quantification of the prediction uncertainty, exploring the safety margin by introducing realistic dispersivity coefficients, and proper illustration of some results relevant for providing feedback to the category A program.

Secondary objective g (6,9)
Quantify prediction uncertainty
The uncertainty related to the groundwater flow and corresponding solute transport simulations should be quantified by inverse conditioning on the available observations. This is not straightforward given a high-dimensional forward model. Therefore, an efficient methodology should be developed for obtaining this uncertainty. Moreover, the same approach should be used for the reference model to quantify the gain of accounting for multi-scale heterogeneity.

Secondary objective h (4,9)
Introduce dispersivity coefficients in transport model
The realistic local-scale and macroscopic dispersivity estimates should be introduced in the solute transport models, in order to quantify its effects more accurately. When considering homogeneous K fields, macroscopic dispersion should be used; when accounting for subsurface heterogeneity through the use of heterogeneous K fields, local-scale dispersion should be used.

Secondary objective i (9)
Provide feedback for the cAt program
The obtained probability distributions of the main solute plume characteristics should be reported for providing feedback to the cAt program.

The obtained flow solution ensembles are however equally important, and should allow for more in-depth understanding of the hydrogeological system.

1.4. Overview of the different chapters

Most of the secondary objectives are addressed in different parts, or by different approaches, in different chapters.
Figure 1.1 illustrates which chapters address the individual secondary objectives and which secondary objectives are represented in each chapter. An overview of the different chapters is given below:

Chapter 2 (a,b,c)[3]
Quantifying outcrop hydraulic conductivity and its spatial variability
Hydraulic conductivity and its spatial variability is characterized with a handheld air permeameter for a small set of outcrops. The obtained data, with a very small measurement support, is upscaled to the outcrop-scale using a numerical approach, yielding a full 2D vertical K tensor. The air permeameter-based predictions are validated as well in this chapter using laboratory measurements.

Chapter 3 (a,b,c)
Derivation of flow and transport parameters from outcropping sediments of the Neogene aquifer, Belgium
The methodology developed in Chapter 2 is applied to a larger set of outcrops in this chapter, and the methodology is extended for estimating mechanical dispersivity from outcrop data. The relation between K variability and dispersivity, as obtained in this chapter, is used in Chapter 9 to predict realistic dispersivity values, in a spatially distributed way, for use in the solute transport simulation.

Chapter 4 (a,b)
Testing outcrop hydrogeological parameters with independent borehole data
The usefulness of outcrop analogue air permeameter measurements for analyzing aquifer heterogeneity is tested in this chapter by performing a comparison between the outcrop-derived parameters obtained in Chapter 2, and

[3] The letters between brackets indicate the respective objectives.

the existing borehole data of the same lithostratigraphical units.

Chapter 5 (a,c,e)
High resolution hydraulic conductivity logging of borehole cores using air permeability measurements
The handheld air permeameter is used on a large set of borehole cores, and the corresponding calibrated high-resolution data is used as a secondary variable for enhanced interpolation of the existing laboratory measurements. This resulted in highly detailed, high-resolution *K* logs, which are used for the calibration of the high-resolution geotechnical direct push data in Chapter 8.

Chapter 6 (e)
Estimation of hydraulic conductivity and its uncertainty from grain-size data using GLUE and artificial neural networks
A new look is given at the classical problem of estimating hydraulic conductivity from grain size by combining artificial neural networks with general likelihood estimation (GLUE). With the developed data-driven modelling tool, the hydraulic conductivity in the center part of the Neogene aquifer at the study area is estimated. These results complement the existing datasets of lab analyses on borehole cores, resulting in a full profile of hydraulic conductivity throughout the Neogene aquifer. The results are incorporated in the updated groundwater flow model, outlined in Chapter 8.

Chapter 7 (a,d)
High-resolution hydrostratigraphic characterization of a heterogeneous sedimentary aquifer using cone penetration data.
This chapter investigates the use of model-based clustering for site-specific soil behaviour type classification of CPT data. Different approaches are compared, and a methodology of automated lithostratigraphical mapping is proposed. The results are compared with a more time-consuming manual approach. One of the tested SBT classification approaches is used in Chapter 8 for enhancing the detection of thin clay lenses from the CPT data.

Chapter 8 (c,e)
Multi-scale aquifer characterization and groundwater flow model parameterization using direct push technologies
Chapter 8 presents the refinement of the groundwater flow model, starting from the reference model with homogeneous layers, and the direct conditioning on all available direct push and borehole data. The hydraulic direct push data is introduced here as well. This chapter makes use of the results obtained in Chapters 5-7. Moreover, some of the assumptions made are supported by observations at the outcrops studied in Chapters 2-4.

Chapter 9 (f,g,h,i)
Large-scale stochastic inversion using non-stationary variograms for transport uncertainty assessment of a near surface radioactive waste disposal
In this chapter, we develop a stochastic approach to obtain an ensemble of flow solutions from the direct conditioned groundwater flow model presented in Chapter 8, which are all consistent with the available observations. For this purpose, we combine the Adaptive Metropolis algorithm and adaptive spatial resampling by block sampling. Both McMC sampling and optimization is performed to quantify the uncertainty on the solute transport simulation used in the category A project. Moreover, the dispersivity results from Chapter 3 are used to obtain a realistic 3D dispersivity distribution.

Chapter 10 provides the general conclusions of this work and recapitulates the main conclusions from Chapter 2-9. Recommendations for future research, and the application of the developed methodologies are provided as well.

References

Anagnostopoulos, A., Koukis, G., Sabatakakis, N., Tsiambaos, G., 2003. Empirical correlations of soil parameters based on Cone Penetration Tests (CPT) for Greek soils. *Geotechnical and Geological Engineering* **21**(4): 377-387.

Bailey, L., Becker, D., Beuth, T., Capouet, M., Cormenzana, J. L., Cuñado, M., Galson, D.A., Griffault, L., Marivoet, J., Serres, C., 2011. Performance Assessment Methodologies in Application to Guide the Development of the Safety Case - European Handbook of the state-of-the-art of safety assessments of geological repositories-Part 1 (p. 266).

Beerten, K., Wemaere, I., Gedeon, M., Labat, S., Rogiers, B., Mallants, D., Salah, S., Leterme, B., 2010. Geological, hydrogeological and

hydrological data for the Dessel disposal site. Project near surface disposal of category A waste at Dessel. STB-SIE(HYD) - Version 1, NIROND-TR 2009-05 E (p. 261).

Blouin, M., Martel, R., Gloaguen, E., 2012. Accounting for Aquifer Heterogeneity from Geological Data to Management Tools. *Ground water* **51**(3): 421-31. doi:10.1111/j.1745-6584.2012.00982.x

Boschan, A., Nœtinger, B., 2012. Scale Dependence of Effective Hydraulic Conductivity Distributions in 3D Heterogeneous Media: A Numerical Study. *Transport in Porous Media* **94**(1): 101-121. doi:10.1007/s11242-012-9991-2

De Barros, F. P. J., Ezzedine, S., Rubin, Y., 2011. Impact of hydrogeological data on measures of uncertainty, site characterization and environmental performance metrics. *Advances in Water Resources* **36**: 51-63. doi:10.1016/j.advwatres.2011.05.004

De Marsily, G., Delay, F., Goncalves, J., Renard, Ph., Teles, V., Violette, S., 2005. Dealing with spatial heterogeneity. *Hydrogeology Journal* **13**(1): 161–183.

Dietrich, P., Leven, C., 2006. Direct push-technologies. In R. Kirsch (Ed.), Groundwater geophysics, A Tool for Hydrogeology (2nd ed., pp. 347–366). Springer. doi:10.1007/978-3-540-88405-7_12

Dietrich, P., Butler, J.J., Faiss, K., 2008. A rapid method for hydraulic profiling in unconsolidated formations. *Ground water* **46**(2): 323–8.

Gedeon, M., Mallants, D., 2004. Hydrogeological model for the safety evaluation . Groundwater flow and transport calculations for the nuclear zone Mol-Dessel. (p. 53).

Gedeon, M., Mallants, D., 2012. Sensitivity Analysis of a Combined Groundwater Flow and Solute Transport Model Using Local-Grid Refinement: A Case Study. *Mathematical Geosciences* **44**(7): 881–899. doi:10.1007/s11004-012-9416-3

Gedeon, M., Mallants, D., Vandersteen, K., Rogiers, B., Laloy, E., 2011. Hydrogeological modelling of the Dessel site - Overview report. Project near surface disposal of category A waste at Dessel. NIROND–TR 2008–15 E V2, June 20 (p. 227).

Hill, M., Kavetski, D., Clark, M., Ye, M., Lu, D., 2012. Uncertainty Quantification for Environmental Models. SIAM News, 45(9).

Huysmans, M., Dassargues, A., 2005. Stochastic analysis of the effect of heterogeneity and fractures on radionuclide transport in a low permeability clay layer, *Enviromental Geology* **48**(7): 920-930.

Huysmans, M., Dassargues, A., 2009. Application of multiple-point geostatistics on modelling groundwater flow and transport in a cross-bedded aquifer (Belgium). *Hydrogeology Journal* **17**(8): 1901–1911.

Jiang, Z., Schrank, C., Mariethoz, G., Cox, M., 2013. Permeability estimation conditioned to geophysical downhole log data in sandstones of the northern Galilee Basin, Queensland: Methods and application. *Journal of Applied Geophysics* **93**: 43-51. doi:10.1016/j.jappgeo.2013.03.008

Köber, R., Hornbruch, G., Leven, C., Tischer, L., Grossmann, J., Dietrich, P., Weiss, H., Dahmke, A., 2009. Evaluation of combined direct-push methods used for aquifer model generation. *Ground Water* **47**(4): 536–46. doi:10.1111/j.1745-6584.2009.00554.x

Koltermann, C., Gorelick, S., 1996. Heterogeneity in Sedimentary Deposits: A Review of Structure Imitating, Process Imitating, and Descriptive Approaches. *Water Resour. Res* **32**(9) : 2617–2658.

Laloy, E., Rogiers, B., Vrugt, J. a., Mallants, D., Jacques, D., 2013. Efficient posterior exploration of a high-dimensional groundwater model from two-stage Markov chain Monte Carlo simulation and polynomial chaos expansion. *Water Resources Research* **49**(5): 2664–2682. doi:10.1002/wrcr.20226

Lessoff, S.C., Schneidewind, U., Leven, C., Blum, P., Dietrich, P., Dagan, G., 2010. Spatial characterization of the hydraulic conductivity using direct-push injection logging. *Water Resources Research* **46**(12): 1–9.

Li, Y., Chan Hilton, A., 2007. Optimal groundwater monitoring design using an ant colony optimization paradigm. *Environmental Modelling & Software* **22**(1): 110–116. doi:10.1016/j.envsoft.2006.05.023

Liu, G., Butler, J.J., Bohling, G.C., Reboulet, E., Knobbe, S., Hyndman, D.W., 2009. A new method for high-resolution characterization of hydraulic conductivity. *Water Resources Research* **45**(8): 1-6.

Lunne, T., Robertson, P.K., Powell, J. J. M., 1997. Cone Penetration Testing in Geotechnical Practice. Blackie Academic and Professional, London, pp. 312

Mariethoz, G., Renard, P., Caers, J., 2010. Bayesian inverse problem and optimization with iterative spatial resampling. *Water Resources Research* **46**(11): 1–17. doi:10.1029/2010WR009274

Meyer, D. W., Tchelepi, H. A., Jenny, P., 2013. A fast simulation method for uncertainty quantification of subsurface flow and transport. *Water Resources Research* **49**(5): 2359-2379. doi:10.1002/wrcr.20240

Nilsson, B., Højberg, A., Refsgaard, J., Troldborg, L. 2007. Uncertainty in geological and hydrogeological data. *Hydrology and Earth System Sciences* **11**: 1551-1561.

ONDRAF/NIRAS. 2010. The cAt-project in Dessel. A long-term solution for Belgian category A waste. Retrieved from http://www.niras-

cat.be/downloads/cAt_masterplanENG.pdf on 07-12-2011.

Quinnan, J., Welty, N., Killenbeck, E., 2010. Hydrostratigraphic and permeability profiling for groundwater remediation projects. In Proceedings of the 2nd International symposium on Cone Penetration Testing (CPT'10).

Reed, P., Minsker, B. S., 2004. Striking the balance: long-term groundwater monitoring design for conflicting objectives. *Journal of Water Resources Planning and Management* **130**(2): 140–149.

Refsgaard, J. C., Christensen, S., Sonnenborg, T. O., Seifert, D., Højberg, A. L., Troldborg, L., 2011. Review of strategies for handling geological uncertainty in groundwater flow and transport modeling. *Advances in Water Resources* **36**: 36-50. doi:10.1016/j.advwatres.2011.04.006

Robertson, P.K. 2010. Estimating in-situ soil permeability from CPT & CPTu. Proceedings of the 2nd International symposium on Cone Penetration Testing (CPT'10), 2-51, 8.

Rojas, R., Kahunde, S., Peeters, L., Batelaan, O., Feyen, L., Dassargues, A., 2010. Application of a multimodel approach to account for conceptual model and scenario uncertainties in groundwater modelling. *Journal of Hydrology* **394**(3-4): 416–435. doi:10.1016/j.jhydrol.2010.09.016

Ronayne, M. J., Gorelick, S.M., Zheng, C., 2010. Geological modeling of submeter scale heterogeneity and its influence on tracer transport in a fluvial aquifer. *Water Resources Research* **46**(10): 1-9.

Rovey, C.W., Cherkauer, D.S., 1995. Scale dependency of hydraulic conductivity measurements. *Ground Water* **33**(5): 769–780.

Schmelzbach, C., Tronicke, J., Dietrich, P., 2011. Three-dimensional hydrostratigraphic models from ground-penetrating radar and direct-push data. *Journal of Hydrology* **398**(3-4): 235–245. doi:10.1016/j.jhydrol.2010.12.023

Sudicky, E. A., Illman, W. A., Goltz, I. K., Adams, J. J., Mclaren, R. G., 2010. Heterogeneity in hydraulic conductivity and its role on the macroscale transport of a solute plume: From measurements to a practical application of stochastic flow and transport theory. *Water Resources* **46**: 1–16. doi:10.1029/2008WR007558

Swartjes, F. A. (Ed.). 2011. Dealing with Contaminated Sites, From Theory Towards Practical Application. Springer. doi:10.1007/978-90-481-9757-6

Tartakovsky, D., Guadagnini, A., Guadagnini, L., 2000. Short Note: Effective Hydraulic Conductivity and Transmissivity for Heterogeneous Aquifers. *Mathematical geology* **32**(6): 751–759.

Tillmann, A., Englert, A., Nyari, Z., Fejes, I., Vanderborght, J., Vereecken, H., 2008. Characterization of subsoil heterogeneity, estimation of grain size distribution and hydraulic conductivity at the Krauthausen test site using Cone Penetration Test. *Journal of contaminant hydrology* **95**(1-2): 57-75.

Van Der Wal, T., Goedemoed, S., Peuchen, J., 2010. Bias reduction on CPT-based correlations. Proceedings of the 2nd International symposium on Cone Penetration Testing (CPT'10), 2-26, 7.

Vienken, T., Dietrich, P., 2011. Field evaluation of methods for determining hydraulic conductivity from grain size data. *Journal of Hydrology* **400**(1-2): 58–71. doi:10.1016/j.jhydrol.2011.01.022

Yang, Y., Wu, J., Sun, X., Wu, J., Zheng, C., 2013. A niched Pareto tabu search for multi-objective optimal design of groundwater remediation systems. *Journal of Hydrology* **490**(20): 56-73. doi:10.1016/j.jhydrol.2013.03.022

Zlotnik, V.A., Zurbuchen, B.R., Ptak, T., Teutsch, G., 2000. Support volume and scale effect in hydraulic con- ductivity: experimental aspects. In: Theory, modelling, and field investigations in hydrogeology: A special volume in honor of Shlomo P Neuman's 60th birthday. Geological Society of America, pp 215–231.

Chapter 2

Quantifying outcrop hydraulic conductivity and its spatial variability

Based on Rogiers B, Beerten K, Smeekens T, Mallants D, Gedeon M, Huysmans M, Batelaan O, Dassargues A. 2013. The usefulness of outcrop analogue air permeameter measurements for analysing aquifer heterogeneity: Quantifying outcrop hydraulic conductivity and its spatial variability. Hydrological processes. DOI: 10.1002/hyp.10007

Abstract

Saturated hydraulic conductivity (*K*) is one of the most important parameters determining groundwater flow and contaminant transport in both unsaturated and saturated porous media. While several well-established laboratory methods exist for determining *K*, *in situ* measurements of this parameter remain very complex and scale-dependent. Often the limited accessibility of subsurface sediments for sampling means an additional impediment to our ability to quantify subsurface *K* heterogeneity. One potential solution is the use of outcrops as analogues for subsurface sediments.

This paper investigates the use of air permeameter measurements on outcrops of unconsolidated sediments to quantify *K* and its spatial heterogeneity on a broad range of sediment types. The Neogene aquifer in northern Belgium is used as a case study for this purpose. To characterize the variability in *K*, 511 small-scale air permeability measurements were performed on outcrop sediments representative over five of the aquifer's lithostratigraphic units. From these measurements, outcrop-scale equivalent *K* tensors were calculated using numerical upscaling techniques. Validation of the air permeameter-based *K* values by comparison with laboratory constant head *K* measurements reveals a correlation of 0.93. Overall the results indicate that hand-held air permeameters are very efficient and accurate tools to characterize saturated *K*, as well as its small-scale variability and anisotropy on a broad range of unconsolidated sediments. The studied outcrops further provided a qualitative understanding of aquifer hydrostratigraphy and quantitative estimates about *K* variability at the centimetre- to meter-scale.

2.1. Introduction

To move beyond traditional methods of aquifer characterization, outcrops may serve as analogues of aquifer sediments to quantify small-scale spatial variability (centimeter- to meter-scale) of subsurface hydrodynamic properties. This becomes particularly relevant if limited accessibility of aquifer sediments hampers collection of core samples suitable for determining material properties such as saturated hydraulic conductivity (*K*). Such knowledge is important since small-scale variability in *K* has implications at larger scales at which many practical applications regarding groundwater flow and contaminant transport operate (Koltermann and Gorelick, 1996; De Marsily *et al.*, 2005; Huysmans and Dassargues, 2009; Ronayne *et al.*, 2010). Moreover, existing inverse methods that incorporate such spatial variability (Hassan *et al.* 2008; Mariethoz *et al.* 2010; Rubin *et al.* 2010) to derive parameter distributions are computationally expensive. It is often not feasible to use such inverse methods for large-scale groundwater modelling, though significant progress has been made (Keating *et al.* 2010; Winton *et al.* 2011; Laloy *et al.* 2013). Stochastic inversion of aquifer structure using sparse geological, geophysical and hydrogeological information was demonstrated by Harp *et al.* (2008). Complex hydrogeological structures are however not normally revealed by modelling, but usually provided as prior information, requiring extensive knowledge on aquifer sediments or depositional processes

(Huysmans *et al.* 2008; Possemiers *et al.* 2012; Michael *et al.* 2010; Edington and Poeter, 2006).

Water flux-based methods to determine small-scale variability in *K* require expensive coring, and are time consuming for low permeability sediments (*e.g.* Yu *et al.*, 2012). Other methods to determine *K* such as indirect texture-based methods lack accuracy, unless site-specific calibration is performed, again requiring direct measurements (*e.g.* Vienken and Dietrich, 2011; Rogiers *et al.*, 2012a; Chapter 6). Considerable advance has been made in the field of Direct Push technologies, with novel hydraulic tools allowing for efficient high-resolution *K* profiling (Dietrich and Leven, 2009; Lessoff *et al.* 2010; Vienken *et al.* 2012a, b). Nonetheless, quantifying small-scale three-dimensional spatial variability or linking *K* profiles to sedimentary structures remains resource intensive.

A very convenient and cost-effective alternative to laboratory-based quantification of *K* spatial variability in the field is the hand-held air permeameter. Since the 1950's, it is recognized that measurements of air permeability (k_a) are more rapid, non-destructive to sediment structure, and pose fewer practical problems than measurements of *K* based on water flux methods (Kirkham, 1947). Consequently, k_a measurements on soils and sediments are increasingly being used to study subsurface hydrodynamic properties (*e.g.* Huysmans *et al.* 2008; Possemiers *et al.* 2012). Probe permeameters are mostly made of a compressed-gas or vacuum cylinder through which gas can be released into or subtracted from porous media. Leakage between the cylinder outlet and the porous medium flow-through area is avoided by placing a ring of compressible, impermeable material at the probe tip. Gas flow rate and pressure are monitored and can be transformed into gas permeability using the analytical equation proposed by Goggin *et al.* (1988a):

$$k_a = \frac{2\mu P_1 Q_1}{\left((P_1^2 - P_2^2) G_0 \mathrm{a}\right)} \qquad [2.1]$$

where k_a is air permeability (m^2); P_1 and P_2 are respectively injection and outflow pressure (Pa); μ is gas viscosity at atmospheric pressure (Pa s); Q_1 is the volumetric airflow rate (m^3/s); G_0 is the Goggin dimensionless geometric factor and *a* is the radius of the seal aperture (m).

Different devices for air permeametry have been developed and described in literature (Chandler *et al.*, 1989; Davis *et al.*, 1994; New England Research and Vindum Engineering, 2011), resulting in very cost-effective ways to perform several hundreds of *in situ* k_a measurements a day, typically on natural outcrops (Goggin *et al.* 1988b; Goss and Zlotnik, 2007), quarries (Huysmans *et al.*, 2008; Possemiers *et al.* 2012), or soil pits (Iversen *et al.* 2003; Beerten *et al.*, 2012). k_a measurements have further been used to validate grain-size-based *K* estimates (Eggleston and Rojstaczer, 2001). More recently, both large- and small-scale anisotropy were studied in more detail (Huysmans *et al.* 2008; Chief *et al.* 2008; Possemiers *et al.* 2012), and consequences for k_a measurements and associated correction factors were determined.

Several transfer functions are available to convert k_a to *K* values. This involves linking air and water permeability, and making use of the relation between *K* and intrinsic saturated water permeability $k_{w,s}$:

$$K = k_{w,s} \frac{\rho_w g}{\eta_w} \qquad [2.2]$$

where ρ is density of water (998.23 kg m^{-3} at 20°C), g is the gravitational acceleration (9.81 m/s^2), and η is the dynamic viscosity of water (10^{-3} Pa s at 20°C). In the ideal case, $k_{w,s}$ should equal k_a, and one could substitute $k_{w,s}$ by k_a in Equation 2.2. In reality, totally dry conditions are hard to obtain, and might cause breakdown of soil or sediment structure. Furthermore, the polar characteristics of water might lead to problems, and air at atmospheric pressure does not act as a true fluid continuum in soil (Iversen *et al.*, 2003). Despite these limitations, the commonly assumed interchangeability between k_a and $k_{w,s}$ (disregarding differences due to 'slippage' of gas at the interface with solids) allows measurements of k_a to be used as an indirect method to predict *K* (Loll *et al.*, 1999; Iversen *et al.*, 2003).

Potential disadvantages of k_a measurements are the sensitivity to partial or full saturation of porous media, and sample surface effects such as irregularities and weathering (Huysmans *et al.*, 2008). Therefore, one must pay attention to the sealing, and prepare a preferably dry and planar outcrop face. Chief *et al.* (2008) showed that errors induced by anisotropic conditions at the scale of the measurement volume can go up

to a factor 2, especially for surface permeameters with large diameter/height ratios.

The importance of spatial heterogeneity is dependent on the scale of heterogeneity compared to that of groundwater flow. Although small-scale variability might have important consequences on contaminant transport, the influence on regional groundwater flow will be limited. Therefore, hydrogeological models generally do not incorporate sub-meter-scale K spatial variability for making predictions at scales relevant to groundwater management. For large-scale models, k_a-based K values can therefore not be used directly. A single equivalent K tensor can however be calculated, representative of the entire grid. This enables K to be quantified, as well as vertical anisotropy at larger scales than that of a single measurement, which is more informative on the properties of hydrogeological units at regional scale.

Several approaches exist to upscale K values. Vereecken *et al.* (2007) defined upscaling as the process that replaces a heterogeneous domain with a homogeneous one so that both domains produce the same response for upscaled boundary conditions. They define the flow parameters in the upscaled homogeneous domain as 'effective' parameters, whereas Renard and de Marsily (1997) prefer the term 'equivalent'. The simplest method consists of using arithmetic and harmonic means of local values (*e.g.* Fleckenstein and Fogg, 2008; Rogiers *et al.*, 2010a). This approach is however only suited for infinitely continuous layered media. Stochastic approaches based on two-point geostatistics can be useful when a spatial model of the heterogeneity is available, and the full small-scale heterogeneous dataset is not known (*e.g.* Flach *et al.*, 2005). For upscaling based on such statistical parameters, the term 'effective' is more appropriate according to Renard and de Marsily (1997). If data for the small scale are available, but derivation of reliable statistical properties is impossible, numerical approaches can be used to obtain a K tensor that satisfies certain criteria, such as flow conservation or head reproduction (*e.g.* Zhou *et al.*, 2010; Li *et al.*, 2011). In the remainder of this paper the term equivalent values will be used to indicate such calculated K tensors.

The objectives of this paper are the following: (i) to demonstrate reliability of hand-held air permeameter measurements in determining small-scale variability of unconsolidated sediments present in outcrops, and (ii) to demonstrate an approach to convert such small-scale hand-held air permeameter measurements into equivalent outcrop-scale K tensors (*i.e.* horizontal hydraulic conductivity K_h, vertical hydraulic conductivity K_v, and potentially off-diagonal components) with corresponding vertical anisotropy (K_h/K_v). For this purpose k_a measurements were performed on five outcrops representing different lithostratigraphical units of the Neogene aquifer in Belgium. Laboratory K values measured on undisturbed samples obtained from different outcropping formations of the Neogene aquifer were used for validation of the air permeameter-based method.

2.2. Hydrogeological setting

The Neogene aquifer is located in the Campine area, northeast Belgium (see Figure 2.1A and B). This aquifer underlies both the Scheldt and Meuse river basins and is considered the most important groundwater reservoir in Flanders (drinking water production exceeds 120×10^6 m³/yr according to VMM, 2005).

The Oligocene (Rupelian) and Mio-Pliocene geology of the study area is presented in Figure 2.1C. Lithology consists of fine to medium grained, glauconitic, micaceous sands with Fe sandstone layers and lignite layers. A varying clay content is found in certain units (*e.g.* Kasterlee, Poederlee, Lillo, Diest Formations (Fms.)), while basal gravels are present between the units (Laga *et al.* 2001). Deposition took place in a shallow marine to perimarine environment at the southern margin of the North Sea Basin (Louwye *et al.*, 2007; Louwye and Laga, 2008; Louwye and De Schepper, 2010). Quaternary deposits with various textures and thicknesses unconformably overlie the Neogene units and constitute the upper few meters of the aquifer system. The Boom Clay aquitard forms the lower boundary of the system (see *e.g.* Yu *et al.*, 2012).

Stresses on groundwater quantity and quality are mainly caused by groundwater abstraction, pollution by diffuse sources including agriculture, former metallurgic industry (Seuntjens *et al.*, 2002), coal mines and various point sources (Coetsiers *et al.*, 2008). Groundwater extraction is mostly from deep wells in the Formations of Diest and Berchem to avoid use of lower quality shallow groundwater.

Notwithstanding the lithological differences between the different formations, Patyn *et al.* (1989) concluded from hydrogeological

assessments that all sediments above the Boom Clay behave as a single aquifer. More recent investigations resulted however in an update of the hydrogeological classification (Meyus *et al.*, 2000; VMM, 2008) and in correspondingly more refined groundwater models (Gedeon *et al.*, 2007; Gedeon, 2008). From 2008 to 2010, the hydrogeological characterization of this aquifer further intensified in the framework of developing a surface repository for low and intermediate level short-lived radioactive waste (ONDRAF/NIRAS, 2010; Beerten *et al.*, 2010; Gedeon *et al.*, 2011; Rogiers *et al.*, 2010a,b, 2011, 2012a (Chapter 6), 2012b).

Five outcrops from three key formations (Mol, Kasterlee and Diest Formation, Table 2.1) representing five different locations within the lithostratigraphy of the Neogene aquifer were selected to demonstrate the proposed methodology (see Figure 2.1C). For a more complete overview on outcropping sediments of the Neogene aquifer, the reader is referred to Rogiers *et al.* (2013a; Chapter 3).

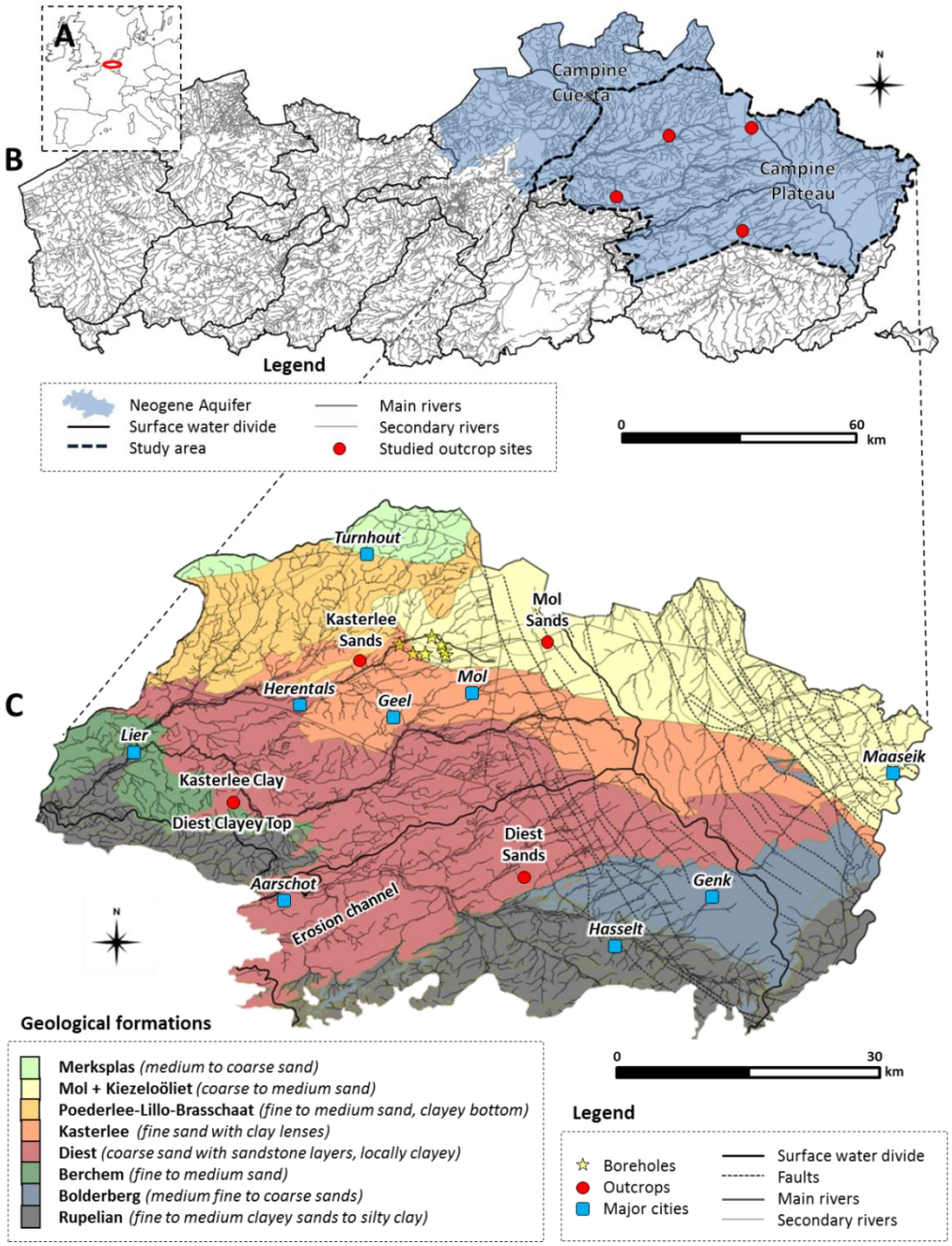

Figure 2.1: Geographical situation of (A) Flanders within Europe, (B) the studied outcrops and the Neogene aquifer within Flanders. (C) Geological map of the Tertiary formations within the studied part of the Neogene aquifer.

Table 2.1: Overview of the five outcrops and the number of *in situ* air permeability measurements (# k_a).

Stratigraphical unit	Lithology	Age	Location	# k_a	Sampling distance
Mol Formation	Sand	Pliocene	Lommel	32	20 cm
Kasterlee Formation	Sand *(sandy part)*	Miocene	Lichtaart	112	10 cm
	Sand with extensive thick clay lenses *(clayey part)*	Miocene	Heist-op-den-Berg	127	10 cm
Diest Formation	Sand with silty/clayey bands *(clayey part)*	Miocene		192	5 cm
	Sand with minor, very thin clay lenses *(sandy part)*	Miocene	Lummen	48	10 cm

2.3. Methods

2.3.1. Field-based air permeameter measurements

We used the Tinyperm II device (New England Research and Vindum Engineering, 2011), which consists of a vacuum cylinder, pressure transducer, handle and plunger, and a microprocessor and control unit. The device is pressed against the outcrop face, and the plunger is depressed to create a vacuum, causing air to flow from the unsaturated porous medium into the device where the gas flow rate and pressure are monitored by the pressure transducer and analysed by the microprocessor unit. Using signal processing algorithms, the unsteady state response function is computed and related to the sample k_a. The exact value of k_a can be determined by an equipment specific calibration curve provided by New England Research. To prevent loose sand debris being sucked into the device, a custom made metallic screen was fitted at the outlet. This required recalibration of the TinyPerm II device (see also Huysmans *et al.*, 2008) to correct for modifications to the air flow. Repeatability of the measurements was tested by one of the operators of the TinyPerm II during this research, on a set of different lithologies with K ranging from $10^{-3.5}$ to $10^{-6.5}$ m/s, with maximum $\log_{10}(K)$ error variance of 0.007.

The volume of sediment involved in a permeameter measurement for isotropic porous media is defined by a hemisphere two to four times the internal radius of the tip seal (Goggin *et al.*, 1988a; Jensen *et al.*, 1994). The TinyPerm II had an inner tip diameter of 9 mm, resulting in an investigation depth of 9-18 mm. Sedimentary structures with a thickness below 9 mm (*e.g.* thin clay drapes) remain difficult to characterize as part of the air flow would bypass the targeted structure. With this device, k_a measurements in permeable sands take a few seconds, less permeable samples take up to a few minutes and clays might take several dozens of minutes.

The empirical equation proposed by Iversen *et al.* (2003) was used in this study:

$$K = 4.87 \times 10^{-10} (k_a)^{1.22} \quad [2.3]$$

with k_a in mD and K in m/s. All values resulting from the air permeameter measurements reported from here on represent saturated K and are obtained through Equation 2.3.

The standard procedure for studying the outcrops consisted of clearing a rectangular outcrop face, marking a regular grid with cells of 20x20 cm with a rope, and performing four measurements at regular distance within each 20x20 cm square. This results in a 10 cm distance between samples, both in horizontal and vertical direction, as obtained for the sandy Kasterlee and Diest Fm. outcrops.

Since the sandy material in the Mol Fm. outcrop was very loosely packed, a sampling density

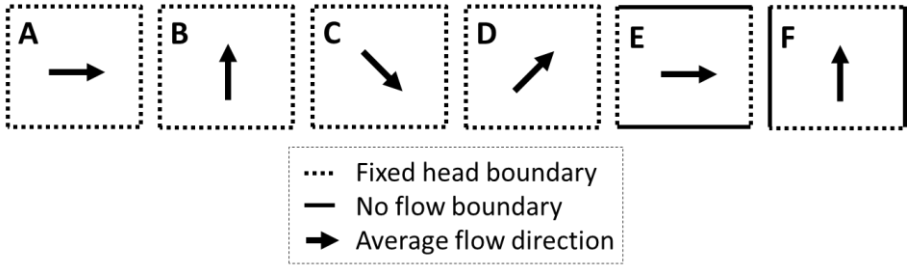

Figure 2.2: Overview of the different boundary condition set-ups.

with 20 cm spacing was applied, preventing collapse of the excavation.

The clayey Diest Fm. outcrop was measured at 5 cm spacing to capture the visibly distinguishable heterogeneity: Khaki-coloured 0.5-cm to a few cm-thick layers consisting of finer textured material, alternating with the more typical Diest Formation sands characterized by a mixture of white and dark green grains and extensive burrowing structures.

As it was impractical to clear the entire 16-m long clayey Kasterlee Fm. outcrop (due to its extent and vegetation), this outcrop was sampled selectively. In part of the outcrop, several measurements were performed each 10 cm in both horizontal and vertical direction, while in the remaining part, four vertical profiles were sampled with 10 cm spacing in the profiles and 2 to 4 m between profiles.

2.3.2 Variography

Variograms are frequently used in relation to structural features in rock successions, to assess the proportion of spatial variability in rock properties including permeability and to identify small-scale heterogeneities (*e.g.* Jensen *et al.*, 1996; McKinley *et al*,. 2004; Corbett and Jensen 1991). Moreover, it is a prerequisite to performing sequential gaussian simulation which has been applied by McKinley *et al.* (2011), conditioned to probe permeability data.

Therefore, we quantify and compare the observed heterogeneity at the different outcrops by performing variography, using the gstat package (Pebesma, 2004). The experimental variograms are calculated directionally, with a tolerance angle of 10°. Spherical models or pure nuggets where initialised manually and fitted automatically using a weighted least squares approach. The weights are represented by the number of data pairs for each experimental variogram point. In cases where the total sill is not reached in the data, the manual estimate using the maximum semivariance is used to avoid obtaining unrealistic parameters.

Both horizontal and vertical variograms are calculated and modelled separately to allow for more flexibility and avoid over-interpretation or models consisting of multiple structures. These independently modelled variograms are easily combined by joining the different structures, and providing the proper anisotropic ranges.

2.3.3. Hydraulic conductivity upscaling

The numerical upscaling approach by Li *et al.* (2011) was adopted for obtaining full two-dimensional K tensors. The main arguments for a numerical approach are the following: (i) despite the large number of measurements, some outcrops are too small to detect spatial correlation and derive meaningful geostatistical models (eliminating geostatistical upscaling approaches); (ii) some outcrops exhibit distinct discrete features (*e.g.* thin clay lenses) and are often highly non-stationary (violation of assumptions behind common geostatistical approaches); and (iii) the advantage of measurements on regular grids allows to transform the data directly to numerical grids for numerical upscaling purposes.

The measurements on the sampling grid were converted into a numerical grid, with one extra grid cell at all sides, called the grid skin. In a first approach, four different flow model setups were defined for each measurement grid, including horizontal flow, vertical flow, and two diagonal flow patterns. In all cases, a fixed head boundary condition was applied to the skin cells, and K values were taken from the adjacent measurements, as shown in Figure 2.2 (A-D).

This approach yields the following overdetermined system of equations:

$$-\begin{bmatrix} \nabla\bar{h}_{x1} & 0 & \nabla\bar{h}_{z1} \\ 0 & \nabla\bar{h}_{z1} & \nabla\bar{h}_{x1} \\ \nabla\bar{h}_{x2} & 0 & \nabla\bar{h}_{z2} \\ 0 & \nabla\bar{h}_{z2} & \nabla\bar{h}_{x2} \\ & \dots & \end{bmatrix}\begin{bmatrix} K_{xx} \\ K_{zz} \\ K_{xz} \end{bmatrix} = \begin{bmatrix} \bar{q}_{x1} \\ \bar{q}_{z1} \\ \bar{q}_{x2} \\ \bar{q}_{z2} \\ \dots \end{bmatrix} \qquad [2.4]$$

where subscripts 1, 2, ... refer to the corresponding boundary conditions, $\nabla\bar{h}_x$ and $\nabla\bar{h}_z$ represent the average horizontal and vertical head gradients (an arbitrary head difference of 1 m across the model domains was used), and K_{xx}, K_{zz}, and $K_{xz} = K_{zx}$ are the K tensor components. Since this system of equations is overdetermined, a least-squares procedure was used to obtain the best-fit equivalent K tensor.

The second approach was added since in many types of sediments the off-diagonal entries in the K tensor can be neglected because the coordinate axes align with the principal axes of the structural heterogeneity. Furthermore, many numerical groundwater flow models only allow for the use of a diagonal K tensor (Li *et al.*, 2011). Therefore, a synthetic constant-head permeameter-type setup was also used for calculating equivalent K values. For this purpose, two additional cases were defined: a horizontal and vertical permeameter-type laboratory setup (Figure 2.2 E-F). The source and sink skin cells were attributed a high K value, and the remaining skin cells were made impermeable. This resulted in the following equations for the diagonal values of the 2-D K tensor:

$$-\begin{bmatrix} \nabla\bar{h}_x & 0 \\ 0 & \nabla\bar{h}_z \end{bmatrix}\begin{bmatrix} K_{xx} \\ K_{zz} \end{bmatrix} = \begin{bmatrix} \bar{q}_x \\ \bar{q}_z \end{bmatrix} \qquad [2.5]$$

Both approaches were compared for all five outcrops to verify (i) if the off-diagonal components are negligible, and (ii) to quantify the influence of the calculation method on the estimated anisotropy.

2.3.4. Constant head lab tests

To validate the k_a-based K values, 26 cylindrical core samples of 100 cm³ (Kopecky rings) were taken from different outcropping formations of the Neogene aquifer. The ring samples were pushed horizontally in the sediment, centered around a k_a measurement location to ensure maximum compatibility. Once the ring was in place, the sediment around was removed to retrieve the sample. Plastic caps were placed on both sides to prevent loss of sediment material and preserve the *in situ* moisture content. All 100 cm³ ring samples were analysed in the lab using a constant-head permeameter (Klute 1965). Because of the very low gradients applied, permeability of the samples was not altered during the tests. Total porosity was also determined for all 26 core samples, and bulk density and moisture content at the time of sampling for 18 out of 26 samples.

2.4. Results and discussion

2.4.1. Outcrop air permeameter measurements

Boxplots of all K values obtained from the k_a measurements are provided in Figure 2.3. As not all sampling schemes were unbiased, for the number of low K values in the clayey Kasterlee Formation outcrop is artificially under-represented due to the longer measurement times, the differences between outcrops should be interpreted with care. Nevertheless, Figure 2.3 provides a reasonable composite picture of the different lithostratigraphic units. The horizontal and vertical directional experimental variograms are presented in Figure 2.4, with the fitted variogram models.

2.4.1.1. Mol Formation

The geometric average K value of the Mol Fm. outcrop (Figure 2.5A) is $10^{-3.7}$ m/s, indicating very permeable material. Spatial correlation seems to be absent, as indicated by the pure nuggets in Figure 2.4. By comparison, Mallants *et al.* (2000) determined a vertical spatial range of 1 m for Mol Sands based on samples from an 8-m long core incorporating K values from above and below the groundwater table. With a data range in K less than one order of magnitude, the overall variability within this outcrop clearly is very small (σ^2 ($\log_{10}K$) = 0.05). Comparing the variograms with the other outcrops (Figure 2.4) clearly shows that this is the most homogeneous outcrop studied. This is due to the high degree of sorting and the maturity of the sand, which was deposited in a continental/estuarine environment.

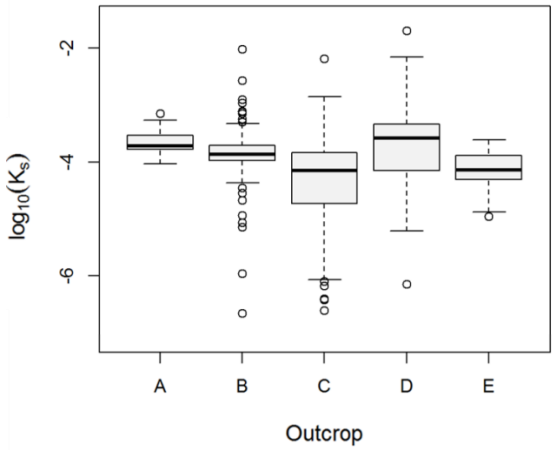

Figure 2.3: Boxplots of the predicted *K* values for all measurements of (A) the Mol Fm., (B) the sandy and (C) the clayey part of the Kasterlee Fm., (D) the clayey and (E) the sandy part of the Diest Fm.

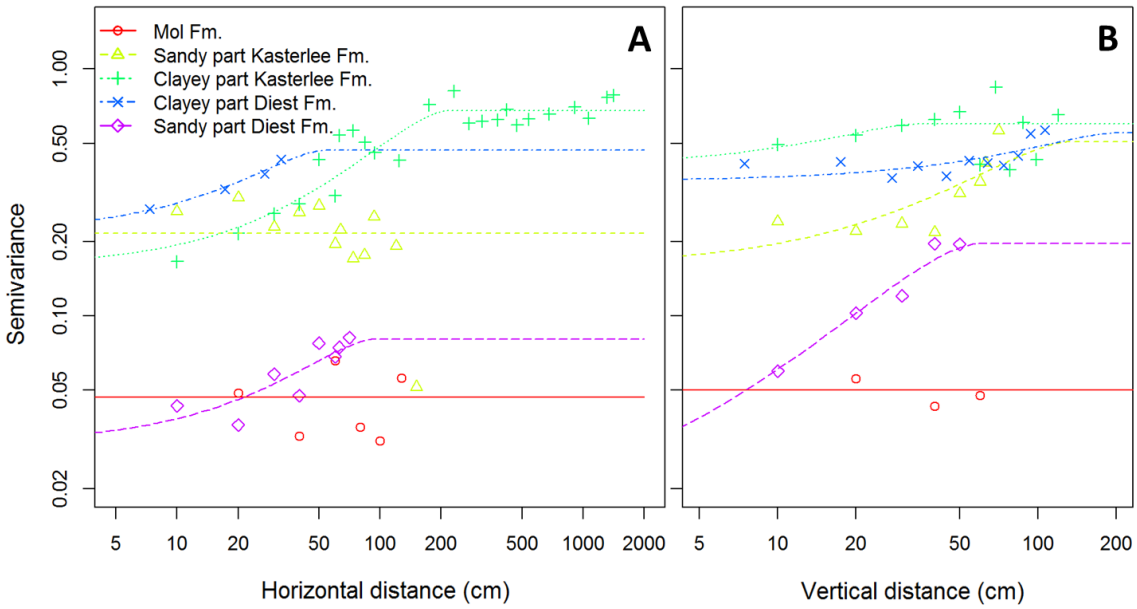

Figure 2.4: Horizontal (A) and vertical (B) experimental and modelled variograms for the different outcrops.

2.4.1.2. Kasterlee Formation: sandy part

The sandy Kasterlee Formation outcrop shows a mean *K* value of $10^{-3.9}$ m/s. While the variance is limited (σ^2 ($\log_{10}K$) = 0.25), several outliers are present (Figure 2.3 and Figure 2.6), with a data range in *K* values of over four orders of magnitude. The 10-cm spaced measurement grid is shown in Figure 2.6A. In the horizontal direction, no spatial correlation seems to exist, similar to the Mol Formation outcrop (Figure 2.4). In the vertical direction however, the semivariance does increase with distance after starting at about the same nugget value. As the total sill is not reached in the experimental variogram, the minimum range is ~1 m. Figure 2.4 also shows that the spatial variability is higher than the Mol and sandy Diest Formations, but lower than the two clayey lithostratigraphical units.

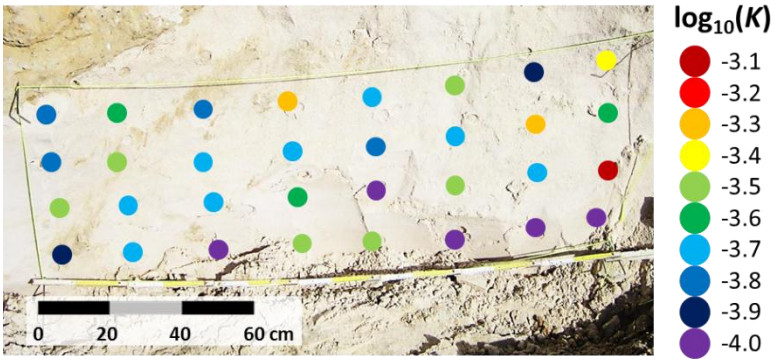

Figure 2.5: The Mol Formation outcrop with air permeameter-based hydraulic conductivity values.

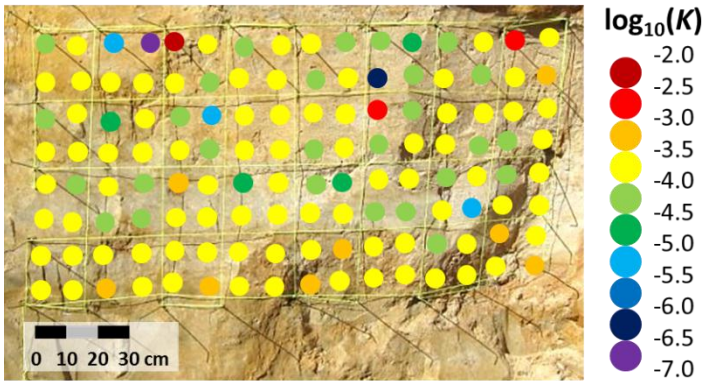

Figure 2.6: The sandy part of the Kasterlee Fm. outcrop with air permeameter-based hydraulic conductivity values.

2.4.1.3. Kasterlee Formation: clayey part

The lowest *K* values in the clayey Kasterlee Fm. outcrop were observed in the clay, ~ $10^{-6.6}$ m/s on average (Figure 2.3; Figure 2.7), which is still relatively high given the high clay content (20%) and low sand content (< 15%). This can be explained by the weathering of the clay, evidenced by penetrating roots and a friable structure. The sand banks separated by these clay lenses exhibited significant differences in *K*. The upper sand bank shows a clay and silt content of respectively 10% and 20%, while *K* ranges from 10^{-4} to $10^{-5.5}$ m/s. The second sand bank has a sand content of more than 90%, and shows higher *K* values, between 10^{-3} and 10^{-4} m/s. The third sand bank is more similar to the first one, both in terms of *K* and grain-size, as is the fourth one. The horizontal experimental variogram shows a steady increase in variability with distance with a spatial range of ~2 m, while the vertical variogram has a much shorter spatial range of ~0.35 m (Figure 2.4). This outcrop has the highest variability at long ranges, but at short ranges the clayey part of the Diest Fm. outcrop has higher variability in the horizontal direction.

2.4.1.4. Diest Formation: clayey part

The visibly distinguishable zones in the clayey Diest Fm. outcrop (Figure 2.8) are clearly the most important features influencing *K*. For the clayey zones *K* ranges between 10^{-4} and 10^{-6} m/s. For the sand in between *K* values are higher than $10^{-4.5}$ m/s, while the coarse sand shows values above $10^{-3.5}$ m/s (Figure 2.3; Figure 2.8B). For the horizontal direction, there seems to be spatial correlation, at least up to ~0.5 m, as the total sill is not reached. One could argue that the vertical direction shows a pure nugget effect, but we chose to fit a spherical model with a high nugget value and a minimum range of ~1 m (Figure 2.4). The spatial variability generally is very high compared to the other outcrops, and at short ranges this outcrop is the most heterogeneous. As the finest material was concentrated in very thin zones, only few

measurements captured them and resulted in K values below 10^{-5} m/s.

2.4.1.5. Diest Formation: sandy part

The outcrop of the sandy part of the Diest Formation shows a mean K value of $10^{-4.1}$ m/s. The 10 cm-spaced grid is shown in Figure 2.9A. The K range exceeds one order of magnitude, making it more heterogeneous than the Mol Formation outcrop, but less than the sands of the Kasterlee Formation (Figure 2.3). The main differences with the fine grained part of the Diest Formation are the pronounced homogeneity of the coarse grained sands and the lower average K value for the sands. Lower K values for coarser grained material compared to finer grained is counter intuitive, but is due to the large heterogeneity and presence of coarse sands in the outcrop of the clayey part of the Diest Formation, and possibly to the physical degradation of the clays resulting in very friable material. Spatial correlation seems to be present both in horizontal (minimum range of ~1 m) and in vertical (range of ~0.5 m) directions (Figure 2.4). Spatial variability is moderate, but there is a high anisotropy. The sampling density, as well as the measurement support did not allow including the very thin clay drapes in the analyses, but their limited horizontal and vertical extent suggests they have little influence. Figure 2.9 suggests that dispersed clay in the upper part of the outcrop might be the reason for the lower K values.

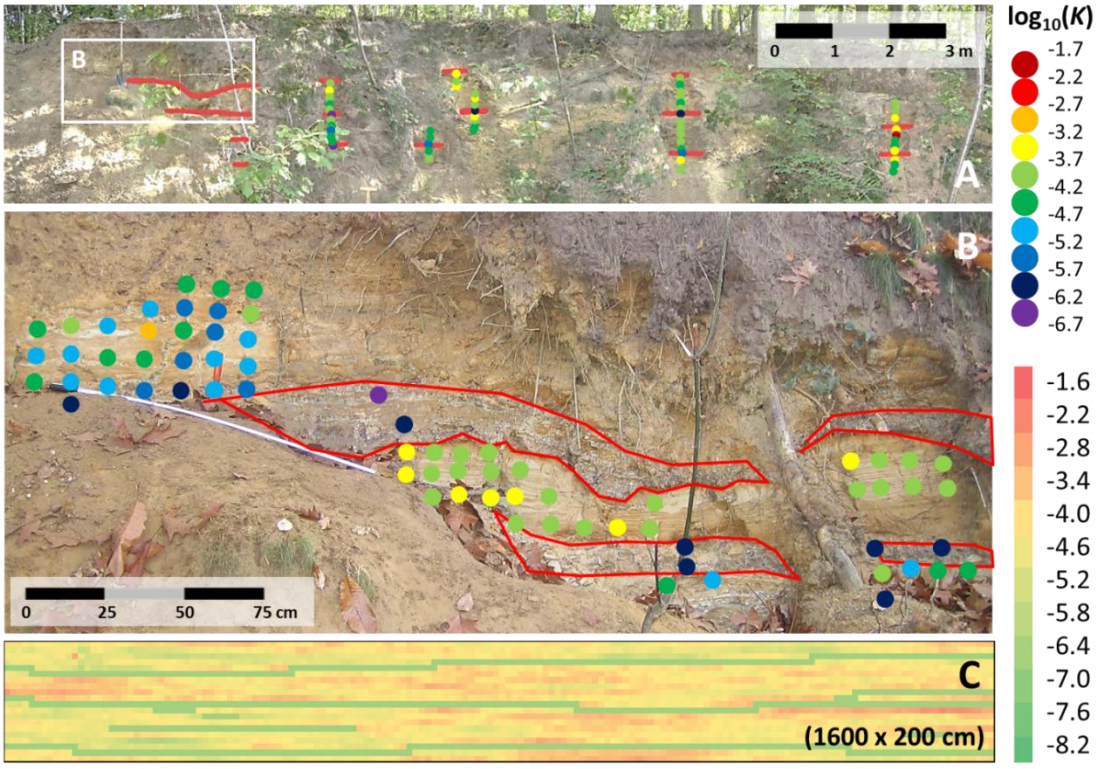

Figure 2.7: (A) The clayey part of the Kasterlee Fm. outcrop with air-permeameter-based hydraulic conductivity values (observed clay lenses are indicated in red). (B) Close up of the left section. (C) Simulated continuous K field with a constant value for the clay lenses.

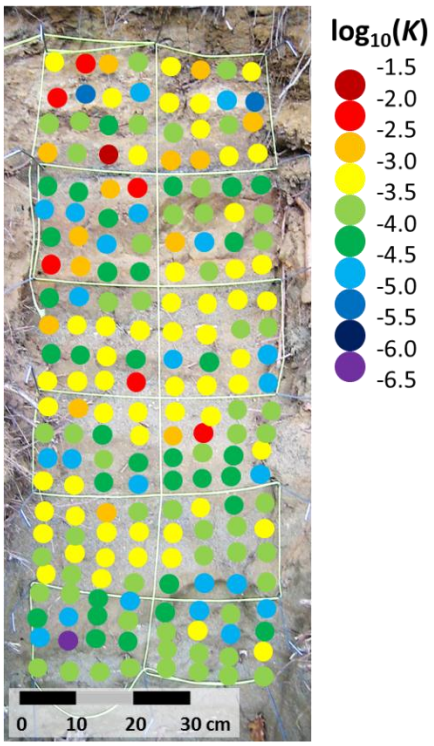

Figure 2.8: The clayey part of the Diest Fm. outcrop with air-permeameter-based *K* values.

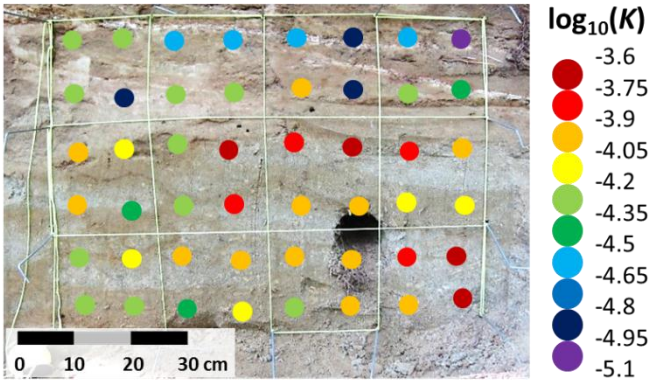

Figure 2.9: The sandy part of the Diest Fm. outcrop with air-permeameter-based *K* values.

2.4.2. Validation of air permeameter *K* estimates

The horizontal *K* values obtained from the 100 cm³ core samples are between 2×10^{-4} and 1.5×10^{-8} m/s (Table 2.2). Bulk density varies between 1.4 and 1.8 g/cm³, water content between 4 and 40%, and total porosity between 36 and 52%. The saturation degree varies between 9 and 85%.

A scatterplot is shown in Figure 2.10. The two outliers are samples taken from an organic layer with high porosity and low bulk density. Although water content measurements were not available for these samples, they were visibly wet during the k_a measurements, hence air flow was seriously hindered and therefore the measurements were biased towards too low values. The remaining samples in Figure 2.10 show a strong linear correlation (r = 0.93) between the constant head measurements and the k_a-based prediction using the Iversen *et al.* (2003) equation. The bias is relatively small, with all predictions more or less within the correct order of magnitude. Beerten *et al.* (2012) used a similar sampling approach in a soil-oriented study.

Table 2.2: Air permeability (k_a), estimated hydraulic conductivity (K_{field}), laboratory-based K measurements (K_{lab}) and other physical properties for the core samples taken from different outcropping formations of the Neogene aquifer.

Nr.	Formation	Bulk density (gr/cm³)	Porosity (cm³/cm³)	K_{lab} (m/s)	k_a (mD)	K_{field} (m/s)	Water content (cm³/cm³)	Saturation degree (%)
1	Kleine Nete alluvium	1.38	0.44	1.02E-04	2.11E+04	9.19E-05	-	-
2		1.41	0.41	2.14E-04	5.38E+04	2.87E-04	-	-
3		1.39	0.45	2.11E-04	4.69E+04	2.43E-04	-	-
4		0.55	0.79	8.89E-07	1.31E+00	6.72E-10	-	-
5		0.35	0.86	2.38E-07	3.14E+00	1.96E-09	-	-
6	Weelde Formation	1.60	0.37	7.60E-05	1.29E+04	5.03E-05	0.04	0.11
7		1.63	0.39	2.03E-05	1.89E+04	8.03E-05	0.17	0.43
8		1.59	0.45	1.25E-06	1.74E+03	4.36E-06	0.38	0.85
9		1.57	0.40	2.03E-04	3.02E+04	1.42E-04	0.04	0.09
10		1.80	0.36	1.95E-08	4.63E+01	5.23E-08	0.30	0.84
11		1.60	0.41	1.29E-05	7.65E+03	2.66E-05	0.21	0.52
12	Sterksel Formation	1.47	0.44	1.90E-04	3.97E+04	1.98E-04	0.09	0.20
13		1.49	0.44	2.81E-04	2.42E+04	1.09E-04	0.04	0.10
14		1.51	0.41	1.07E-05	5.82E+03	1.90E-05	0.27	0.66
15	Mol Formation	1.55	0.41	2.88E-04	9.06E+04	5.42E-04	0.04	0.10
16	Poederlee Formation	1.47	0.42	1.14E-04	8.54E+03	3.04E-05	-	-
17		1.59	0.46	2.36E-05	5.64E+02	1.10E-06	-	-
18		1.47	0.43	1.46E-08	6.27E+01	7.57E-08	-	-
19		1.47	0.44	9.66E-05	2.17E+04	9.50E-05	0.06	0.13
20		1.39	0.42	7.00E-05	2.23E+04	9.82E-05	0.09	0.20
21		1.60	0.43	2.11E-05	4.42E+03	1.36E-05	0.23	0.55
22		1.58	0.45	7.85E-05	2.23E+04	9.82E-05	0.13	0.28
23		1.44	0.52	2.27E-07	1.51E+02	2.21E-07	0.41	0.79
24	Kasterlee Formation	1.54	0.45	8.07E-05	3.97E+04	1.98E-04	0.07	0.16
25		1.53	0.45	9.65E-05	2.63E+04	1.20E-04	0.05	0.11
26		1.46	0.43	9.09E-05	9.31E+04	5.61E-04	0.08	0.20

Table 2.3: Overview of equivalent K values and anisotropy factor VANI for the different outcrop grids.

Outcrop	-------Permeameter set-up--------			---Combination of four boundary conditions---			
	$\log_{10}(K_{xx})$	$\log_{10}(K_{zz})$	VANI	$\log_{10}(K_{xx})$	$\log_{10}(K_{zz})$	$\log_{10}(K_{xz})$	VANI
Mol Fm.	-3.64	-3.60	0.92	-3.68	-3.69	-5.37	1.03
Kasterlee Fm: sandy	-3.86	-3.85	0.98	-3.86	-3.87	-5.62	1.02
Kasterlee Fm: clayey	-4.13	-5.71	38.31	-4.05	-5.47	-13.08	26.46
Diest Fm: clayey	-3.56	-3.88	2.08	-3.58	-3.77	-4.96	1.53
Diest Fm: sandy	-4.06	-4.14	1.21	-4.09	-4.18	-12.23	1.22

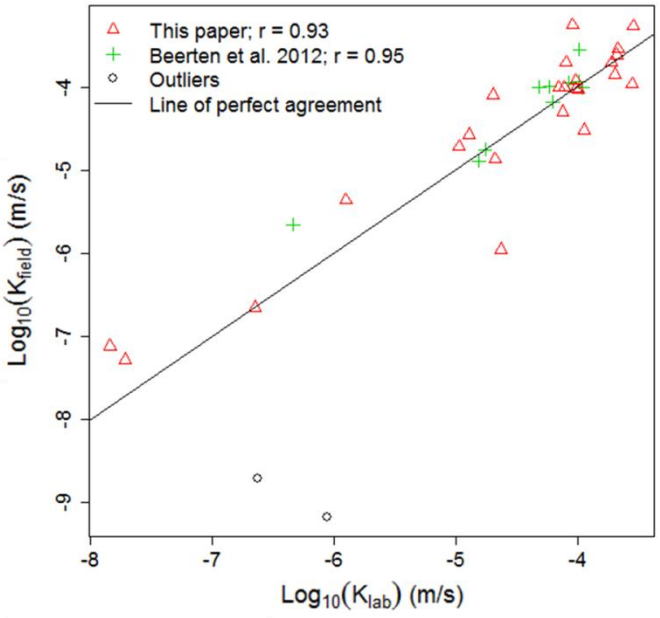

Figure 2.10: Lab hydraulic conductivity values (K_{lab}) *versus* air permeability-based *K* values from the outcrops (K_{field}; obtained with the empirical equation of Iversen *et al.*, 2003), with an overall linear correlation coefficient of 0.93 (without outliers). Results from Beerten *et al.* (2012) are based on the same permeameter and 100 cm³ cores for the lab values for similar outcrops in the same region.

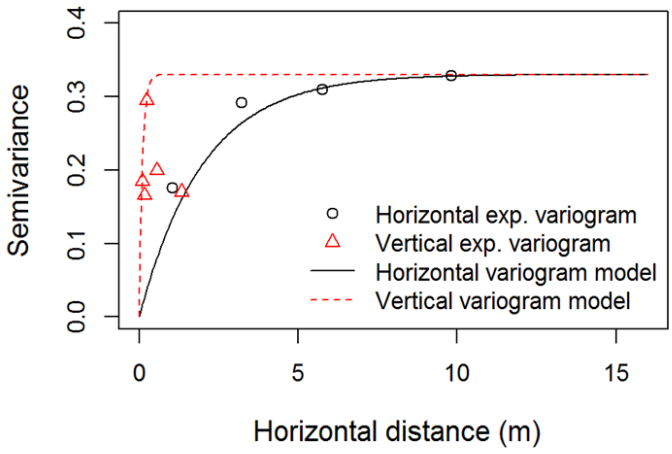

Figure 2.11: Experimental and model variograms for the sands in the clayey Kasterlee Fm. outcrop. For the vertical variogram experimental data beyond lag distances of ~1 m (two data points) are not used in the model fit because of bias due to the limited vertical extent of the outcrop.

Their data are included in Figure 2.10 and are clearly compatible with our observations. We however extend the validated *K* range with 2 orders of magnitude.

These results demonstrate that reliable measurements can be obtained with a saturation degree up to 80%. Since the water content within the outcrop is expected to be higher further away from the air-exposed face, the first few centimetres of the outcrop face probably are much dryer than the 100 cm³ core-based depth-average values. Therefore, for reliable measurements, the saturation degree within the measurement volume should probably be well below 80%. This is also indicated by Dane *et al.* (2011), who demonstrated a saturation degree of

20% reduces k_a by one order of magnitude in comparison with the fully unsaturated case. Given the good agreement with the independent measurements, most of our samples probably had a much lower saturation degree in the sediment volume characterized by the air permeameter, and the saturation degree should preferably be below 20% for reasonable estimates of *K*.

The accuracy of the air permeameter within the *K* range of 10^{-8} to 10^{-3} m/s is hereby verified. This coverage of 5 orders of magnitude is suited for many unconsolidated sediment types.

2.4.3. Outcrop-scale equivalent *K* values

Due to the fact that k_a-based *K* values cannot be used directly in large-scale models, a single *K* tensor was calculated for the entire grid of each outcrop. In doing so, we quantify *K* as well as its vertical anisotropy at a scale several orders of magnitude larger than that of a single measurement. Unlike all other outcrops, the number of sampling locations at the clayey Kasterlee Formation outcrop was too small to generate continuous *K* maps based on measurements alone (Figure 2.7A and B). Sequential Gaussian simulation was therefore performed within the sand layers, while a constant value was used for the clay lenses for which the structure was taken from the field observations. This is justified by the weathered state of the clay which made it difficult to obtain a reasonable number of reliable field measurements. An exponential variogram model was fitted to the *K* measurements within the sand banks (Figure 2.11), with a sill of 0.33 and range of 2 and 0.1 m, respectively in the horizontal and vertical direction. The horizontal correlation length is ~10 m whereas in the vertical direction it is a few tens of cm. This is the result of the limited thickness of the sand layers (between 40 and 70 cm) with few samples (at most 14) taken in the vertical direction. The overall lowest measured *K* value was used for the clay lenses. The results of this simulation are shown in Figure 2.7C.

Equivalent *K* values were then determined for all outcrop *K* grids applying the upscaling techniques described above in Processing Modflow (Chiang and Kinzelbach, 2001). The equivalent *K* tensor values obtained are listed in Table 2.3. The non-zero components of the *K* tensor are given each time, together with the vertical anisotropy factor VANI = K_{xx}/K_{zz}.

Using the permeameter setup, the horizontal equivalent *K* values for all five outcrops are between $10^{-3.56}$ and $10^{-4.13}$ m/s. Similar results are obtained when fixed-head boundary conditions are used (case A – D in Figure 2.2). The overall variability in horizontal equivalent *K* thus remains small. The vertical equivalent *K* values show more variation, ranging from $10^{-3.60}$ to $10^{-5.71}$ m/s, based on the permeameter setup. Again, the four combined fixed head boundary conditions give similar results as the permeameter-type setups. The off-diagonal components of the *K* tensor are all at least one order of magnitude below the diagonal components, which justifies using diagonal-tensor models like *e.g.* MODFLOW (Harbaugh, 2005) in large-scale simulations of the Neogene aquifer (*e.g.* Gedeon *et al.*, 2011; Gedeon and Mallants, 2012; Rogiers *et al.*, 2012b).

It should be noted here that the derived equivalent 2D *K* tensors are only representative for 3D *K*, in which the third dimension is constant. As this is not the case in reality, a discrepancy between the derived 2D and the true 3D values can be expected. For isotropic media, using the variance of $\log_{10}K$ to correct for the discrepancy between 2D and 3D using the results of Gutjahr *et al.* (1978) would lead to an increase in K by at most 11%. Given the lack of clear correlation lengths in all different cases, we can only make a rough estimation for anisotropic media using the generalization of Matheron's conjecture (Matheron 1967) by Ababou (1996). With a horizontal correlation length of 2 m and a vertical one of 0.5 m, which seems reasonable at the outcrop scale given the empirical semivariograms in Figure 2.4, the max discrepancy between 2D and 3D would amount 16% for K_v. For larger anisotropies, the differences only decrease.

Concerning the vertical anisotropy parameter, *VANI*, the Mol and Kasterlee Formation sands seem isotropic at the studied scale, while for the Diest Formation outcrops the *VANI* parameter amounts respectively to 1.2 and 2 (permeameter setup). The highest value (38) is obtained for the clayey Kasterlee Fm. outcrop, which is caused by the alternating clay and sand layers. By comparison, Gedeon and Mallants (2012) obtained an aquifer-scale anisotropy value for the clayey Kasterlee Formation aquitard of 148 based on inverse modelling. The difference is approximately a factor of 3.9, and can be attributed to the degraded state of the outcrop clay lenses.

The obtained outcrop-scale equivalent *K* values suggest that the heterogeneity in the studied sediments translates mainly to differences in the vertical *K* component, while the variability of the horizontal *K* component between the different lithostratigraphical units is limited. Consequently, vertical anisotropy varies between 1 and 38, and should be taken into account for most units, when modelling at the outcrop-scale or larger scales. This suggests that outcrop equivalent *K* values can provide useful qualitative and quantitative insights in *K* values required at more regional groundwater flow modelling scales.

2.5. Summary and conclusions

The handheld air permeameter is an efficient and accurate tool to characterize *K* and its spatial patterns *in situ* on outcrop sediments. It allows quantification of spatial variability in *K* on scales that are not normally feasible by classical core-based techniques. Moreover, the demonstrated approach to obtain equivalent outcrop-scale *K* tensors proved to be very informative and provided *K* values at spatial scales commensurate with that of numerical model grids. Such information can be valuable as prior information for use in geostatistical models, quantification of relative differences between lithostratigraphic units as well as anisotropy for groundwater flow modelling.

We selected five lithostratigraphic layers pertaining to a sandy aquifer and determined *K* on outcrops by means of air permeametry on high-resolution sampling grids. k_a-based mean *K* values ranged from $10^{-4.37}$ to $10^{-3.68}$ m/s. The individual measurements ranged from $10^{-9.17}$ to $10^{-1.56}$ m/s. The obtained *K* patterns clearly represented the sediment texture variations observed in the field, and *K* ranges ranged from less than one order of magnitude to more than four orders of magnitude in the presence of clay lenses.

Validation of the k_a-based *K* estimates was demonstrated by comparison with constant head lab tests on 100 cm³ cores. The results show that the *K* estimates are accurate, at least between 10^{-8} to 10^{-3} m/s.

Concerning spatial heterogeneity, most formations exhibited weak spatial correlation with ranges of ~0.5 m to ~2 m in the horizontal direction. Some outcrops clearly consisted of a single heterogeneous material suited for application of two-point geostatistics, while others, with texturally different materials often alternating over short distances, might require other approaches at this scale, such as multiple-point statistics.

The variation of the upscaled *K* values between the studied outcrops is less than an order of magnitude for horizontal and up to two orders of magnitude for vertical *K*. The calculated vertical anisotropy has a value of at most two for the sandy sediments while it is 38 for the clayey Kasterlee Fm. outcrop. This high value is supported by results from inverse groundwater flow modelling, suggesting that outcrop equivalent *K* values can provide useful insight in aquitard *K* values.

Current limitations of the study include the limited size and small number of the studied outcrops. Increasing both would result in better representativity of the obtained results. Also, there might be a discrepancy between 2D and 3D derived equivalent *K* tensors, which is estimated not to be larger than about 16% for the outcrops studied here. The transferability of the obtained *K* statistics to the corresponding subsurface sediments should be scrutinized before considering them representative of the *in situ* aquifer or aquitard properties. A first attempt to gain some insights into this transferability is presented by Rogiers *et al.* (2013b; Chapter 4).

References

Ababou, R., 1996. Random porous media flow on large 3D grids: numerics, performance, and application to homogenization, Chap 1, vol 79. In: Wheeler, M.F. (ed.) Environmental studies, mathematical, computational & statistical analysis. Springer, New York, pp 1–25

Beerten, K., Wemaere, I., Gedeon, M., Labat, S., Rogiers, B., Mallants, D., Salah, S., Leterme, B., 2010. Geological, hydrogeological and hydrological data for the Dessel disposal site. Project near surface disposal of category A waste at Dessel – Version 1.- Brussels, Belgium: NIRAS/ONDRAF, 2010.- 273 p.- NIROND–TR 2009–05 E V1.

Beerten, K., Deforce, K., Mallants, D., 2012. Landscape evolution and changes in soil hydraulic properties at the decadal, centennial and millennial scale: A case study from the Campine area, northern Belgium. *Catena* **95**: 73-84.

Chandler, M.A., Goggin, D.J., Lake, L.W., 1989. A mechanical field permeameter for making rapid, non-destructive permeability measurements. *Journal of Sedimentary Petrology* **59**(4): 613-635.

Chiang, W., Kinzelbach, W., 2001. 3D-groundwater modeling with PMWIN. Springer: Berlin. ISBN 3-540- 67744-5

Chief, K., Ferré, T.P.A., Hinnell, A.C., 2008. The Effects of Anisotropy on In Situ Air Permeability Measurements. *Vadose Zone Journal* **7**(3): 941.

Coetsiers, M., Blaser, P., Martens, K., Walraevens, K., 2008. Natural background levels and threshold values for groundwater in fluvial Pleistocene and Tertiary marine aquifers in Flanders, Belgium. *Environmental Geology* **57**(5): 1155-1168.

Corbett, P.W.M., Jensen, J.L., 1991. A comparison of small permeability measurement methods for reservoir characterization: Advances in Reservoir Technology: Edinburgh, Petroleum Science and Technology Institute, Conference Paper, 19 p.

Dane, J.H., Vrugt, J.A., Unsal, E., 2011. Soil Hydraulic Functions Determined from Measurements of Air Permeability, Capillary Modeling, and High-Dimensional Parameter Estimation. *Vadose Zone Journal* **10**(1): 459.

Davis, J.M., Wilson, J.L., Phillips, F.M., 1994. A Portable Air-Minipermeameter for Rapid In Situ Field Measurements. *Ground Water* **32**(2): 258–266.

De Marsily, G., Delay, F., Goncalves, J., Renard, P., Teles, V., Violette, S., 2005. Dealing with spatial heterogeneity. *Hydrogeology Journal* **13**(1): 161–183.

Dietrich, P., Leven, C., 2009. Direct push-technologies. In Groundwater geophysics, a tool for hydrogeology, Kirsch R (Ed.). Springer, New York. 2nd ed.; 347-366.

Edington, D., Poeter, E., 2006. Stratigraphic Control of Flow and Transport Characteristics. *Ground Water* **44**(6): 826-831.

Eggleston, J., Rojstaczer, S., 2001. The value of grain-size hydraulic conductivity estimates: Comparison with high resolution in-situ field hydraulic conductivity. *Geophysical Research Letters* **28**(22): 4255.

Flach, G.P., Harris, M.K., Smits, A.D., Syms, F.H., 2005. Modeling aquifer heterogeneity using cone penetration testing data and stochastic upscaling methods. *Environmental Geosciences* **12**(1): 1-15.

Fleckenstein, J.H., Fogg, G.E., 2008. Efficient upscaling of hydraulic conductivity in heterogeneous alluvial aquifers. *Hydrogeology Journal* **16**(7): 1239-1250.

Gedeon, M., Wemaere, I., Marivoet, J., 2007. Regional groundwater model of north-east Belgium. *Journal of Hydrology* **335**(1-2): 133-139.

Gedeon, M., 2008. Neogene Aquifer Model. SCK•CEN Report ER-48: Mol, Belgium; 110. ISSN: 1782-2335.

Gedeon, M., Mallants, D., Vandersteen, K., Rogiers, B., 2011. Hydrogeological modelling of the Dessel site - Overview report. Project near surface disposal of category A waste at Dessel. NIROND-TR report 2008–15 E V2.

Gedeon, M., Mallants, D., 2012. Sensitivity analysis of a combined groundwater flow and transport model. *Mathematical Geosciences* **44**(7): 881-889. DOI: 10.1007/s11004-012-9416-3.

Goss, D., Zlotnik, V.A., 2007. Air permeameter investigation of surficial dune structures in the Nebraska Sand Hills. *AAPG Bulletin* **91**(5): 645–652.

Goggin, D.J., Thrasher, R.L., Lake, L.W., 1988a. A theoretical and experimental analysis of minipermeameter response including gas slippage and high velocity flow effects. *In Situ* **12**: 79–116.

Goggin, D.J., Chandler, M.A., Korcurek, G.A., Lake, L.W., 1988b. Patterns of permeability in eolian deposits: Page Sandstone (Jurassic), northeastern Arizona. *SPE Formation Eval*. **3**: 297-306.

Gutjahr, A.L., Gelhar, L.W., Bakr, A.A., McMillan, J.R., 1978. Stochastic analysis of spatial variability in subsurface flow 2: Evaluation and application. *Water Resour. Res*. **14**: 953–959

Harbaugh, A.W., 2005. MODFLOW-2005, The U.S. Geological Survey modular ground-water model—the Ground-Water Flow Process. U.S. Geological Survey Techniques and Methods 6-A16, USGS.

Harp, D., Dai, Z., Wolfsberg, A., Vrugt, J., Robinson, B., Vesselinov, V., 2008. Aquifer structure identification using stochastic inversion. Geophysical *Research Letters* **35**(8). , DOI: 10.1029/2008GL033585

Hassan, A., Bekhit, H., Chapman, J., 2008. Uncertainty assessment of a stochastic groundwater flow model using GLUE analysis. *Journal of Hydrology* **362**(1-2): 89-109.

Huysmans, M., Peeters, L., Moermans, G., Dassargues, A., 2008. Relating small-scale sedimentary structures and permeability in a cross-bedded aquifer. *Journal of Hydrology* **361**(1-2): 41-51.

Huysmans, M., Dassargues, A., 2009. Application of multiple-point geostatistics on modelling groundwater flow and transport in a cross-bedded aquifer (Belgium). *Hydrogeology Journal* **17**(8): 1901–1911.

Iversen, B.V., Moldrup, P., Schjonning, P., Jacobsen, O.H., 2003. Field Application of a Portable Air Permeameter to Characterize Spatial Variability in Air and Water Permeability. *Vadose Zone Journal* **2**(4): 618-626.

Jensen, J.L., Glasbey, C.A., Corbett, P.W.M., 1994. On the interaction of geology, measurement, and statistical-analysis of small-scale permeability measurements. *Terra Nova* **6**(4): 397–403.

Jensen, J.L., Corbett, P.W.M., Pickup, G.E., Ringrose, P.S., 1996. Permeability semivariograms, geological structure and flow performance. *Mathematical Geology* **28**: 419–435.

Keating, E.H., Doherty, J., Vrugt, J.A., Kang, Q., 2010. Optimization and uncertainty assessment of strongly nonlinear groundwater models with high parameter dimensionality. *Water Resources Research* **46**(10): 1-18.

Klute, A., 1965. Laboratory measurements of hydraulic conductivity of saturated soil. In: Methods of soil analysis. Part 1. Black CA. (ed.). *Agronomy* **9**: 210-220.

Kirkham, D., 1947. Field method for determination of air permeability of soil in its undisturbed state. *Soil Sci. Soc. Am. Proc*. **11**:93–99.

Koltermann, C., Gorelick, S., 1996. Heterogeneity in Sedimentary Deposits: A Review of Structure Imitating, Process Imitating, and Descriptive Approaches. *Water Resources Research* **32**(9): 2617-2658.

Laga, P., Louwye, S., Geets, S., 2001. Paleogene and Neogene lithostratigraphic units (Belgium). *Geologica Belgica* **4**(1-2): 135–152.

Laloy, E., Rogiers, B., Vrugt, J.A., Mallants, D., Jacques, D., 2013 Accelerating Bayesian inference of a highly parameterized and CPU-demanding subsurface flow model using a polynomial chaos expansion to precondition MCMC simulation. *Water Resources Research* **49**(5): 2664–2682, DOI 10.1002/wrcr.20226.

Lessoff, S.C., Schneidewind, U., Leven, C., Blum, P., Dietrich, P., Dagan, G., 2010. Spatial characterization of the hydraulic conductivity using direct-push injection logging. *Water Resources Research* **46**(12): 1–9.

Li, L., Zhou, H., Gómez-Hernández, J.J., 2011. A Comparative Study of Three-Dimensional Hydraulic Conductivity Upscaling at the MAcro-Dispersion Experiment (MADE) site, Columbus Air Force Base, Mississippi (USA). *Journal of Hydrology* **404**(3-4): 278-293.

Loll, P., Moldrup, P., Schjønning, P., Riley, H., 1999. Predicting saturated hydraulic conductivity from air permeability: Application in stochastic water infiltration modeling. *Water Resources Research* **35**(8): 2387–2400.

Louwye, S., De Schepper, S., Laga, P., Vandenberghe, N., 2007. The Upper Miocene of the southern North Sea Basin (northern Belgium): a palaeoenvironmental and stratigraphical reconstruction using dinoflagellate cysts. *Geological Magazine* **144**(1): 33.

Louwye, S., Laga, P., 2008. Dinoflagellate cyst stratigraphy and palaeoenvironment of the marginal marine Middle and Upper Miocene of the eastern Campine area, northern Belgium (southern North Sea Basin). *Geological Journal* **43**(1): 75–94.

Louwye, S., De Schepper, S., 2010. The Miocene–Pliocene hiatus in the southern North Sea Basin (northern Belgium) revealed by dinoflagellate cysts. *Geological Magazine* **147**(05): 760-776.

Mallants, D., Espino, A., Van Hoorick, M., Feyen, J., Vandenberghe, N., Loy, W., 2000. Dispersivity estimates from a tracer experiment in a sandy aquifer. *Ground Water* **38**(2): 304-310.

Mariethoz, G., Renard, P., Caers, J., 2010. Bayesian inverse problem and optimization with iterative spatial resampling. *Water Resources Research* **46**(11): 1-17.

Matheron, G., 1967. Éléments pour une théorie des milieux poreux. Masson and Cie, Paris

McKinley, J.M., Lloyd, C.D., Ruffell, A.H., 2004. Use of variography in permeability characterisation of visually homogeneous sandstone reservoirs with examples from outcrop studies. *Mathematical Geology* **36**: 761–779.

McKinley, J.M., Atkinson, P.M., Lloyd, C.D., Ruffell, A.H., Worden, R.H., 2011. How Porosity and Permeability Vary Spatially With Grain Size, Sorting, Cement Volume, and Mineral Dissolution In Fluvial Triassic Sandstones: The Value of Geostatistics and Local Regression. *Journal of Sedimentary Research* **81**: 844–858.

Meyus, Y., De Smet, D., De Smedt, F., Walraevens, K., Batelaan, O., Van Camp, M., 2000. Hydrogeological coding of the subsurface of Flanders (HCOV). @WEL 8, Water, p. 1-13.

Michael, H.A., Li, H., Boucher, A., Sun, T., Caers, J., Gorelick, S.M., 2010. Combining geologic-process models and geostatistics for conditional simulation of 3-D subsurface heterogeneity. *Water Resources Research* **46**(5): 1-20.

New England Research and Vindum Engineering. 2011. TinyPerm II Portable Air Permeameter, User's Manual. Retrieved from http://www.vindum.com/TinyPermManual.pdf on 14-06-2011.

ONDRAF/NIRAS. 2010. Het cAt-project in Dessel. Een langetermijnoplossing voor het Belgische categorie A-afval. Retrieved from http://www.niras-cat.be/downloads/cAt_masterplan_NL_LOW.pdf on 07-12-2011.

Patyn, J., Ledoux, E., Bonne, A., 1989. Geohydrological research in relation to radioactive waste disposal in an argillaceous formation. *Journal of Hydrology* **109**: 267–285.

Pebesma, E.J., 2004. Multivariable geostatistics in S: the gstat package. *Computers & Geosciences* **30**: 683-691.

Possemiers, M., Huysmans, M., Peeters, L., Batelaan, O., Dassargues, A., 2012. Relationship between sedimentary features and permeability at different scales in the Brussels Sands. *Geologica Belgica* **15**(3): 156-164.

Renard, P., de Marsily, G., 1997. Calculating equivalent permeability: a review. *Advances in Water Resources* **20**(5-6): 253–278.

Rogiers, B., Schiltz, M., Beerten, K., Gedeon, M., Mallants, D., Batelaan, O., Dassargues, A., Huysmans, M., 2010a. Groundwater model

parameter identification using a combination of cone-penetration tests and borehole data. Proceedings of the IAHR International Groundwater Symposium 2010, Valencia, 22-24 September 2010, 19 pp.

Rogiers, B., Mallants, D., Batelaan, O., Gedeon, M., Huysmans, M., Dassargues, A., 2010b. Exploratory and structural data analysis of a sandy aquifer at Mol/Dessel, Belgium. *Geophysical Research Abstracts* **12**: EGU2010-6890-1.

Rogiers, B., Beerten, K., Smeekens, T., Mallants, D., 2011. Air permeability measurements on Neogene and Quaternary sediments from the Campine area: using outcrop analogues for determining hydrodynamic aquifer properties. SCK•CEN Report ER-177: Mol, Belgium; 23.

Rogiers, B., Mallants, D., Batelaan, O., Gedeon, M., Huysmans, M., Dassargues, A., 2012a. Estimation of hydraulic conductivity and its uncertainty from grain-size data using GLUE and artificial neural networks. *Mathematical Geosciences* **44**(6): 739-763.

Rogiers, B., Mallants, D., Batelaan, O., Gedeon, M., Huysmans, M., Dassargues, A., 2012b. The usefulness of CPTs for deterministic spatially heterogeneous, large-scale aquitard parametrisation. In: Models - Repositories of Knowledge, Oswald, S.E., Kolditz, O., Attinger, S. (eds). IAHS Publ. 355, Proceedings ModelCare 2011, Leipzig, Germany, 18-22 September 2011, pp. 41-47. ISBN 978-1-907161-34-6.

Rogiers, B., Beerten, K., Smeekens, T., Mallants, D., Gedeon, M., Huysmans, M., Batelaan, O., Dassargues, A., 2013a. Derivation of flow and transport parameters from outcropping sediments of the Neogene aquifer, Belgium. *Geologica Belgica* **16**(3): 129-147.

Rogiers, B., Beerten, K., Smeekens, T., Mallants, D., Gedeon, M., Huysmans, M., Batelaan, O., Dassargues, A., 2013b. The usefulness of outcrop analogue air permeameter measurements for analysing aquifer heterogeneity: testing outcrop hydrogeological parameters with independent borehole data. *Hydrol. Earth Syst. Sci. Discuss.* **10**: 9689-9720.

Ronayne, M.J., Gorelick, S.M., Zheng, C., 2010. Geological modeling of submeter scale heterogeneity and its influence on tracer transport in a fluvial aquifer. *Water Resources Research* **46**(10): 1-9.

Rubin, Y., Chen, X., Murakami, H., Hahn, M., 2010. A Bayesian approach for inverse modeling, data assimilation, and conditional simulation of spatial random fields. *Water Resources Research* **46**(10): 1-23.

Seuntjens, P., Mallants, D., Simunek, J., Patyn, J., Jacques, D., 2002. Sensitivity analysis of physical and chemical properties affecting field-scale cadmium transport in a heterogeneous soil profile. *Journal of Hydrology* **264**(1-4): 185-202.

Vereecken, H., Kasteel, R., Vanderborght, J., Harter, T., 2007. Upscaling Hydraulic Properties and Soil Water Flow Processes in Heterogeneous Soils: A Review. *Vadose Zone Journal* **6**(1): 1-28.

Vienken, T., Dietrich, P., 2011. Field evaluation of methods for determining hydraulic conductivity from grain size data. *Journal of Hydrology* **400**(1-2): 58-71.

Vienken, T., Leven, C., Dietrich, P., 2012a. Use of CPT and other direct push methods for (hydro-) stratigraphic aquifer characterization: a field study. *Canadian Geotechnical Journal* **49**(2): 197-206.

Vienken, T., Tinter, M., Rogiers, B., Leven, C., Dietrich, P., 2012b. Evaluation of field methods for vertical high resolution aquifer characterization. AGU Fall Meeting, Abstract ID 1494473; San Francisco, USA, 3-7 December 2012.

VMM., 2005. Grondwaterbeheer in Vlaanderen: het onzichtbare water doorgrond. Bijlage 4: Vergunde debieten per grondwaterlichaam. Retrieved from http://www.vmm.be/ on 21-02-2012.

VMM., 2008. Het Centraal Kempisch Systeem, 108.

Winton, C., Pettway, J., Kelley, C.T., Howington, S., Eslinger, O.J., 2011. Application of Proper Orthogonal Decomposition (POD) to inverse problems in saturated groundwater flow. *Advances in Water Resources* **34**(12): 1519-1526.

Yu, L., Rogiers, B., Gedeon, M., Wemaere, I., Marivoet, J., De Craen, M., Mallants, D., 2012. A critical review of laboratory and in-situ hydraulic conductivity measurements for the Boom Clay in Belgium. *Applied Clay Science* **75-76**: 1-12.

Zhou, H., Li, L., Gómez-Hernández, J.J., 2010. Three-dimensional hydraulic conductivity upscaling in groundwater modeling. *Computers & Geosciences* **36**(10): 1224-1235.

Chapter 3

Derivation of flow and transport parameters from outcropping sediments of the Neogene Aquifer, Belgium

*Based on Rogiers B, Beerten K, Smeekens T, Mallants D, Gedeon M, Huysmans M, Batelaan O, Dassargues A. 2013. Derivation of flow and transport parameters from outcropping sediments of the Neogene aquifer, Belgium. Geologica Belgica **16**(3): 129-147.*

Abstract

Centimetre-scale saturated hydraulic conductivities (*K*) are derived from air permeability measurements on a selection of outcrops of the Neogene aquifer in the Campine area, Belgium. Outcrop sediments are of Miocene to Quaternary age and have a marine to continental origin. Grain size analyses for the same outcrops and corresponding *K* predictions using previously developed models are also presented. We discuss outcrop hydrogeological properties and quantify the heterogeneity within the outcrops in detail using geostatistical variography. Moreover, outcrop-scale *K* values, their anisotropy and dispersivities are numerically calculated as a means to upscale such small-scale measurements to a larger scale commensurate with the scale of flow and transport modelling. By studying the small-scale variability as observed in outcrops, we gain crucial understanding of the larger-scale behaviour of the corresponding hydrogeological units within the Neogene aquifer, the most important groundwater reservoir of Flanders. The results of this study will equally improve conceptual hydrogeological model building and parameterization.

3.1. Introduction

The Neogene aquifer underlying the Campine area in northeast Belgium (Fig. 3.1) is considered to be the most important groundwater reservoir in Flanders, with drinking water production exceeding 120×10^6 m³/yr (VMM, 2005). It underlies both the Scheldt and Meuse river basins and consists of Oligocene (Rupelian) to Mio-Pliocene sediments with a Quaternary cover (Fig. 3.2). The main lithologies include fine to medium grained, glauconitic, micaceous sands with Fe sandstone layers and lignite layers. A varying clay content is found in certain units (*e.g.* Kasterlee, Poederlee, Lillo and Diest Formations), while basal gravels are present between the units (Laga *et al.*, 2001). Deposition took place in a shallow marine to perimarine environment at the southern margin of the North Sea Basin (Louwye *et al.*, 2007; Louwye & Laga, 2008; Louwye & De Schepper, 2010). The Quaternary deposits have various textures and thicknesses and unconformably overlie the Neogene units and constitute the upper part of the aquifer system (not shown on Fig. 3.2); the Boom Clay aquitard forms the lower boundary of the aquifer (see *e.g.* Yu *et al.*, 2012).

Groundwater recharge mainly occurs on the elevated regions of the Campine Plateau and the cuesta of the Campine Complex (Fig. 3.1), and the numerous smaller interfluvia distributed across the study area. Groundwater abstraction, pollution by diffuse sources including agriculture, former metallurgic industry (Seuntjens *et al.*, 2002), coal mines and point sources are the main stresses on both groundwater quantity and quality (Coetsiers *et al.*, 2008). Hydrodynamic evolution of the aquifer since 1833 was studied by Van Camp *et al.* (2010), and groundwater chemistry was investigated by Coetsiers & Walraevens (2006). Groundwater abstraction happens mostly via deep wells in the Formations of Diest and Berchem to avoid use of lower quality shallow

groundwater polluted by surface activities; it represents about 12% of the estimated aquifer recharge (Van Camp *et al.*, 2012).

Patyn *et al.* (1989) concluded from hydrogeological assessments that all sediments above the Boom Clay behave as a single aquifer, despite the lithological differences between the different aquifer units. More recent investigations however resulted in an update of the hydrogeological classification (Meyus *et al.*, 2000; VMM, 2008) and in correspondingly more refined groundwater models (Gedeon *et al.*, 2007; Gedeon, 2008; Gedeon & Mallants, 2012). Model updating involved abandoning the concept of a single aquifer in favour of using two aquifers separated by a retarding layer: the Kasterlee Clay aquitard. From 2008-2010, the hydrogeological characterization of the different units within this aquifer has been further intensified in the framework of developing a

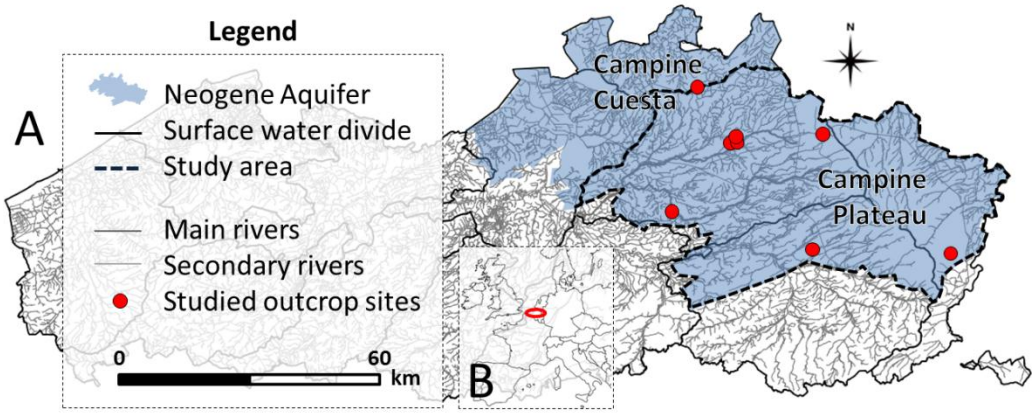

Figure 3.1. A: Location of the outcrop sites and the study area within the Neogene aquifer, in Flanders, Belgium. B: Location of Flanders within Europe.

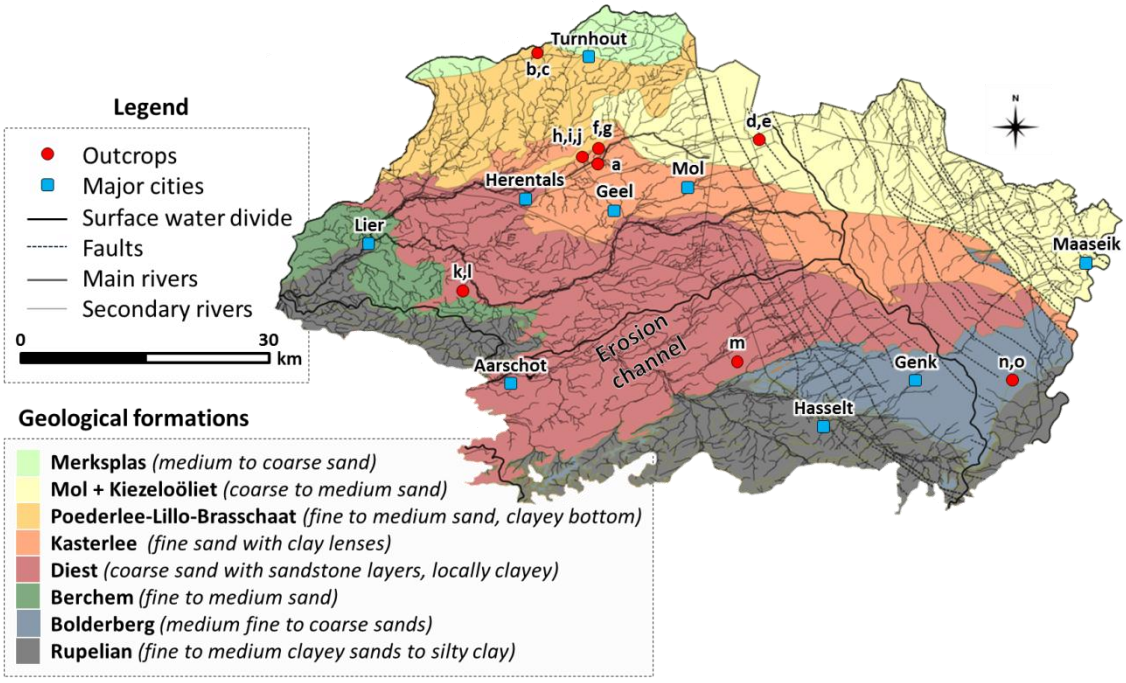

Figure 3.2. Geological map of the Tertiary formations within the studied part of the Neogene aquifer (see study area in Figure 3.1), with the selected outcrop sites (see Table 3.1) and several major cities.

Table 3.1. Overview of the studied outcrops, number (#) of measurements (k_{air}: in-situ air permeability; *gs*: grain size), and upscaling approach.

ID	Location	Formation	Stratigraphy	Age	# k_{air}	# gs	Upscaling
a	Kasterlee	Schelde Group	Kleine Nete deposits	Quaternary	34	5	-
b	Beerse	Weelde	Campine Clay-Sand Complex		58	27	Arithmetic and Harmonic means
c					80	-	Numerical upscaling using flow models and different boundary condition setups
d	Lommel	Sterksel	Lommel Sands		182	-	
e		Mol	Mol Sands	Pliocene	32	-	
f	Kasterlee	Poederlee	Poederlee Sands		84	2	
g					40	1	
h	Lichtaart				130	-	
i			Poederlee-Kasterlee transition zone		29	12	
j		Kasterlee	Kasterlee Sands	Miocene	112	-	
k	Heist-op-den-berg		Kasterlee Clay		127	7	
l		Diest	Diest Clayey Top		192	4	
m	Lummen		Diest Sands		48	-	
n	Maasmechelen	Bolderberg	Bolderberg Sands		80	-	
o					64	-	
				Total=	**1292**	**58**	

surface repository for low and intermediate level short-lived radioactive waste (ONDRAF/NIRAS, 2010).

The investigations included a few hundred cone penetration tests (Rogiers *et al.*, 2010a-c), eight cored boreholes in the Mol/Dessel region and numerous lab analyses (Beerten *et al.*, 2010). This new information resulted in refining the hydrogeological models (Gedeon *et al.*, 2011; Gedeon & Mallants, 2012; Rogiers *et al.*, 2010a) and included quantifying the influence of small-scale (centimetre- to meter-scale) variability within the Neogene Aquifer (Rogiers *et al.*, 2010b-d, 2011, 2012 (Chapter 6), 2013a (Chapter 2)). Upscaling of such small-scale heterogeneity is key to building successful models at the larger, more practical scales typical of groundwater management models (Koltermann & Gorelick, 1996; De Marsily *et al.*, 2005; Huysmans & Dassargues, 2009; Ronayne *et al.*, 2010).

Using outcrop sediments as analogues for aquifer sediments is especially useful for characterizing such small-scale spatial variability in hydrodynamic properties (Huysmans *et al.*, 2008; Possemiers *et al.*, 2012; Rogiers *et al.*, 2013a; Chapter 2). Rogiers *et al.* (2013b; Chapter 4) demonstrated the benefit of outcrop-based hydraulic conductivity estimates and their spatial variability for an improved understanding and quantification of an aquifer's heterogeneity in *K*. This paper presents a follow-up analysis of the Rogiers *et al.* (2013a,b; Chapter 2, 4) study and includes additional outcrop sediments while we also quantify a key contaminant transport parameter (*i.e.* dispersivity). To increase knowledge on the small-scale variability within the Neogene aquifer sediments, this paper presents detailed descriptions of 15 outcrops including grain size characteristics and several different hydraulic conductivity estimates. *K*-estimates include those based on i) grain size distributions using

the predictive models of Rogiers *et al.* (2012; Chapter 6), and ii) in-situ air permeameter measurements at the centimetre scale and their numerically calculated outcrop-scale equivalent values following the methodology of Rogiers *et al.* (2013a; Chapter 2). Finally, we estimate the outcrop-scale dispersivity using an equivalent 2D uniform random walk approach (see *e.g.* Salamon *et al.*, 2006).

3.2. Methods

3.2.1. Outcrop selection

The selection of outcrops was driven by two factors: 1) the need for outcrops analogues to those sediments that play a key role in the groundwater flow and transport modelling that underpin the safety assessment of the ONDRAF/NIRAS surface disposal project (ONDRAF/NIRAS, 2010), and 2) the need for natural analogues of engineered earth covers as part of a design study in which understanding the long-term behaviour of such systems is critical (Jacques *et al.*, 2010). For this purpose, eight locations were selected within the study area (Figs 3.1 & 3.2), resulting in a total of 15 outcrops.

3.2.2. Estimation of hydraulic conductivity

3.2.2.1. Prediction from grain size

Grain size analyses were performed on outcrop sediments as a means to compare outcrops with subsurface sediments (see Rogiers *et al.*, 2013b; Chapter 4); a further objective was to provide saturated hydraulic conductivity estimates independent of the air permeameter values. Grain size analyses were performed by laser diffraction using a Malvern Mastersizer (Malvern Instruments Ltd., UK). The method consists of monitoring the amount of reflection and diffraction that is transmitted back from a laser beam directed at the particles. Prior to analysis, 15g of each sample was dried at 60°C during at least 24h and aggregate-forming material like carbonate cements, organic material and Fe-oxides and hydroxides were removed using acidification and oxygenation. Each sample was then divided into 10 sub-samples by a rotary sample splitter to enable repeated measurements on a single sample. All samples were measured at least twice. The final result was based on the average grain size distribution of all sub-samples. Note that particle sizes are expressed as size of an equivalent sphere with an identical diffraction pattern.

Various approaches exist to relate K to grain size data. Most methods use a single grain size parameter and hence omit the information encompassed by the entire grain size distribution. Rogiers *et al.* (2012; Chapter 6) compared two data-driven modelling methods that use the entire grain size distribution data as input for prediction of K, *i.e.* multiple linear regression (MLR) and artificial neural networks (ANNs). Besides the predictive capacity of the methods, the uncertainty associated with the model predictions was also evaluated, since such information is important for stochastic groundwater flow and contaminant transport modelling. The ANN method was therefore combined with a generalized likelihood uncertainty estimation (GLUE) approach to predict K from grain size data. The resulting K-models were shown to be superior to non site-specific literature-based prediction models (Chapter 6), including several of the models discussed by Vienken and Dietrich (2011) and the predictive K model of the Rosetta software (Schaap *et al.*, 2001). The GLUE-ANN ensemble prediction also proved to be slightly better than the MLR model, and the prediction uncertainty was reduced by half an order of magnitude on average. Since the data used to develop the predictive K-models all originated from an area within the current study area, the same models are used in this paper to provide estimates of hydraulic conductivity. Since the superiority of site-specific calibrations was demonstrated (Rogiers *et al.*, 2012; Chapter 6), other non site-specific estimates are omitted. It should be noted however that some of the outcrop sediments display grain size fractions not observed in the dataset used to develop the models. Using the MLR model for such grain sizes would not introduce large errors; the non-linear GLUE-ANN model however might provide erroneous estimates for grain size distributions that were not used during model calibration ("training" of the ANNs).

3.2.2.2. Air permeameter-based measurements of hydraulic conductivity at the cm-scale

Centimetre-scale air permeameter measurements were performed with the Tinyperm II device

(New England Research & Vindum Engineering, 2011). The device had an inner tip diameter of 9 mm, resulting in a depth of investigation between 9 and 18 mm, allowing for a very detailed small-scale analysis. The calibration was adjusted after fitting a filter into the device tip, as presented by Huysmans *et al.* (2008). The transfer function proposed by Iversen *et al.* (2003) was used to derive hydraulic conductivity estimates; validation of the approach based on laboratory-derived *K*-values on similar outcrop core samples was reported by Rogiers *et al.* (2013a; Chapter 2).

3.2.2.3. Upscaling from cm-scale to outcrop-scale

Where it was possible to measure *K* on regular 2D grids on the outcrop faces, the numerical upscaling approach by Rogiers *et al.* (2013a; Chapter 2) and Li *et al.* (2011) was applied to calculate outcrop-scale *K* values. This consists of creating 2D groundwater flow models whose numerical grids correspond to the measurement grids, with a set of artificial boundary conditions, and calculating a single equivalent *K* tensor for the entire domain by using the average fluxes and head gradients. Processing Modflow (Chiang & Kinzelbach, 2001) was used for this purpose. If measurements were made only along a single depth profile, arithmetic and harmonic means were calculated instead to obtain an outcrop-scale representative value, assuming that we are dealing with a perfectly layered medium. In one single case, several measurements were scattered over a large outcrop and geostatistical simulation was used to provide a grid for upscaling to outcrop-scale representative values.

3.2.3. Estimation of outcrop-scale dispersivity

Spreading of solute in groundwater in the direction of groundwater flow (longitudinal dispersion) and transverse to the main flow direction (transverse dispersion) may be calculated on the basis of the longitudinal (α_L) and transverse dispersivity (α_T) and the porewater velocity *v*, *e.g.* $D_L = \alpha_L \times v$. In this paper particle tracking was used to derive α_L and α_T based on the artificial boundary condition setups used to upscale the air permeameter measurements. Fig. 3.3 shows, as an example, the different boundary conditions that were applied to the Kasterlee Sands outcrop (vertical cross-section corresponding to the measurement grid at the outcrop face). Effects of heterogeneity in *K* on calculated heads, flow vectors, and particle pathlines are displayed for the different boundary conditions.

We used an equivalent 2D uniform random walk (Salamon *et al.*, 2006) to describe the arrival times and locations of particles at the downgradient side of both permeameter-type setup simulations; the following equations describe the particle positions in the system (for the horizontal flow case):

$$x_p^k = x_p^{k-1} + v_x \Delta t + \sqrt{2 D_L \Delta t} X_L \qquad [3.1]$$

$$z_p^k = z_p^{k-1} + v_x \Delta t + \sqrt{2 D_T \Delta t} X_T \qquad [3.2]$$

with x_p^k the *x* coordinate of particle *p* at timestep *k*, v_x the average velocity, Δt the time step, D_L and D_T the longitudinal and transversal dispersion coefficients, and X_L and X_T two independent random numbers from a standard normal distribution. Assuming Fickian dispersion, which might not be justified in all cases, reordering these equations, and normalizing by the time to travel to the downgradient side of the flow model, for each individual particle, gives the following normal distributions with standard deviation $\sqrt{2D}$

$$\frac{x_p^k - x_p^{k-1} - v_x \Delta t}{\sqrt{\Delta t}} = \sqrt{2 D_L} X_L \qquad [3.3]$$

$$\frac{z_p^k - z_p^{k-1} - v_x \Delta t}{\sqrt{\Delta t}} = \sqrt{2 D_T} X_T \qquad [3.4]$$

The quantities on the left side can easily be obtained from the forward particle tracking end points. The standard deviations of these are then used to estimate both D_L and D_T, and based on the average velocity v_x the dispersivities α_L and α_T are determined as $\alpha_L = D_L / v_x$ and $\alpha_T = D_T / v_x$.

Using the results of the particle tracking, outcrop-scale longitudinal and transversal dispersivity values can thus be calculated, both for horizontal and vertical flow, using the flow matrices from the *K* upscaling simulations. The estimated dispersivities are estimated at the outcrop scales (at most several m travel distance), for larger travel distances a dispersion-scale relationship has to be invoked (Gelhar *et al.*, 1992; Schulze-Makuch, 2005).

The obtained values are only representative for relatively short travel distances (a few m) in 2D. Braiding of flow lines can occur in 3D, while in 2D they can never cross. Transverse dispersivity therefore falls back to zero in 2D, while longitudinal dispersivity is underestimated in 2D. Beaudoin & de Dreuzy (2013) show numerically that this underestimation can be considerable in case of isotropic multi-Gaussian log-permeability fields. For anisotropic media, the effect is expected to be less (*e.g.* Gelhar 1993). Without clear correlation lengths for all outcrops, it is decided not to perform any correction as this does not imply changes of orders of magnitude.

3.3. Overview of the outcrop results

Outcrop sediments were investigated at 8 different locations, resulting in a total of 15 outcrops covering 11 different geological formations (see overview in Table 2.1). In addition to in-situ air permeability measurements, grain size data were collected for a subset of the sampling locations. An overview of the *K*-transformed air permeability measurements is presented in Fig. 3.4; subsequent sections provide a discussion of the geological formations and their corresponding outcrops on a one by one basis.

Descriptive statistics of all hydraulic conductivity values obtained from air permeameter measurements are given in Table 3.2. Heterogeneity in *K* is discussed on the basis of the mean and standard deviation (SD) of $\log_{10}$-transformed hydraulic conductivity, minimum and maximum values, range (max - min) and the geometric coefficient of variation ($CV=100\times SD/Mean$). The corresponding boxplots are given in Fig. 3.5. Not all sampling schemes were unbiased as the number of very low *K* values typical of fine-textured layers is underrepresented in some cases due to a very long measurement time or high moisture content. Regardless of these limitations of the measurement technique, considering all outcrop data, a reasonable composite picture of the different sediments within the heterogeneous aquifer is obtained.

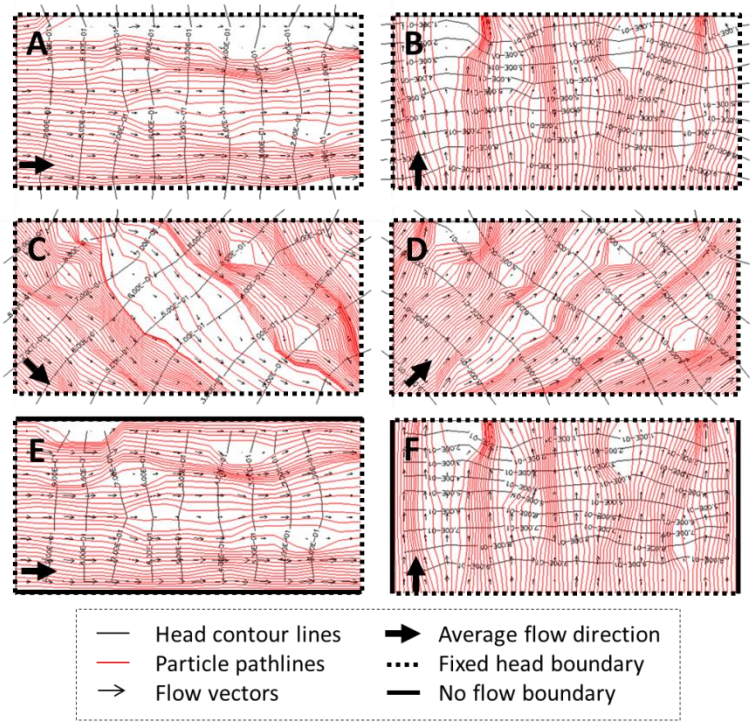

Figure 3.3. Particle tracking simulations for the Kasterlee Sands outcrop. A: horizontal flow setup, B: vertical flow setup, C, D: diagonal flow setup, E, F: horizontal and vertical permeameter-type setup. The modelled vertical cross-section represents the measurement grid at the outcrop face, with one grid cell for each obtained air permeability measurement.

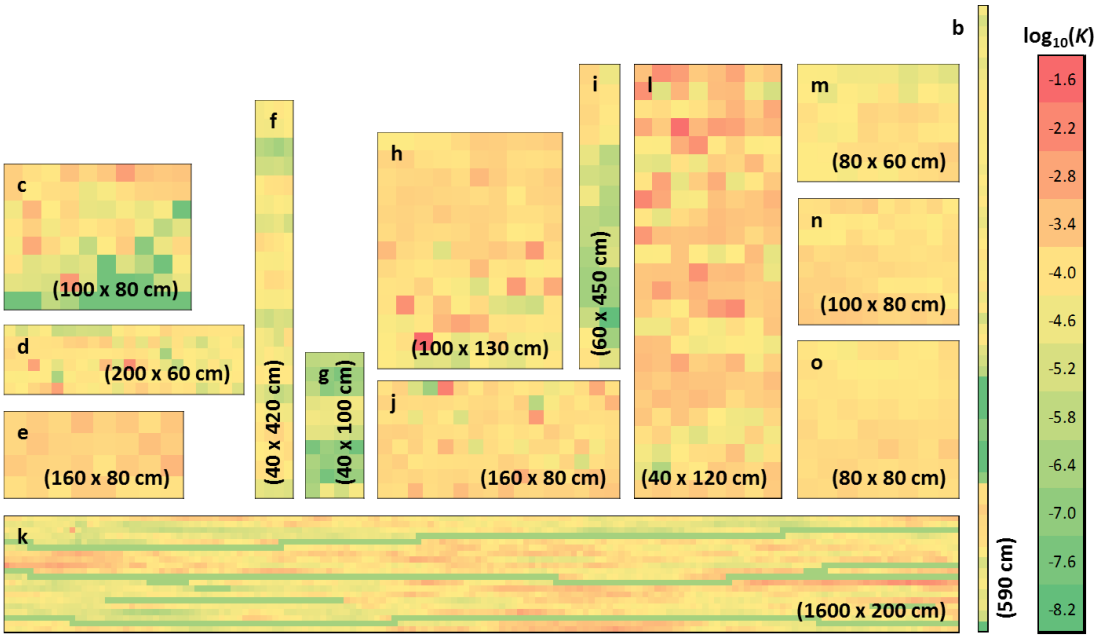

Figure 3.4. Overview of the air-permeability derived *K*-values for all outcrops (except outcrop a); data grids used for numerically upscaling *K*-values. Grid k (Kasterlee Clay) is the only outcrop where geostatistical simulation is used for filling in missing values.

Table 3.2. Extreme values (Min, Max), mean, range, standard deviation (SD), and coefficient of variation (CV) of log_{10}-transformed hydraulic conductivity *K* for all outcrops based on air permeability measurements (k_{air}).

ID		Stratigraphy	# k_{air}	Geomean	---------- $log_{10}(K)$ ----------					
				K (m/s)	Min	Mean	Max	Range	SD	CV
a	point bar sands	Kleine Nete deposits	31	1.88E-04	-4.3	-3.72	-3.4	0.84	0.24	-6%
	channel beds		3	1.39E-09	-9.2	-8.86	-8.3	0.87	0.48	-5%
b		Campine Clay-Sand Complex	58	2.20E-05	-8.1	-4.66	-3.6	4.55	0.92	-20%
c			80	4.00E-05	-7.8	-4.4	-2.5	5.25	1.02	-23%
d		Lommel Sands	182	6.16E-05	-6.4	-4.21	-2.3	4.12	0.59	-14%
e		Mol Sands	32	2.10E-04	-4	-3.68	-3.2	0.89	0.22	-6%
f		Poederlee Sands	84	3.13E-05	-6.5	-4.5	-3.7	2.81	0.61	-14%
g			40	1.15E-06	-7.8	-5.94	-4.4	3.43	1.07	-18%
h			130	1.26E-04	-5.5	-3.9	-1.6	3.88	0.53	-14%
i		Poederlee-Kasterlee transition zone	29	5.20E-06	-8.2	-5.28	-3.7	4.48	1.14	-22%
j		Kasterlee Sands	112	1.33E-04	-6.7	-3.88	-2	4.64	0.5	-13%
k		Kasterlee Clay	127	4.24E-05	-6.6	-4.37	-2.2	4.42	0.82	-19%
l		Diest Clayey Top	192	2.08E-04	-6.2	-3.68	-1.7	4.45	0.64	-17%
m		Diest Sands	48	7.23E-05	-5	-4.14	-3.6	1.34	0.32	-8%
n		Bolderberg Sands	80	1.38E-04	-4.2	-3.86	-3.5	0.77	0.19	-5%
o			64	1.21E-04	-4.3	-3.92	-3.5	0.79	0.15	-4%
Total			***1292***							

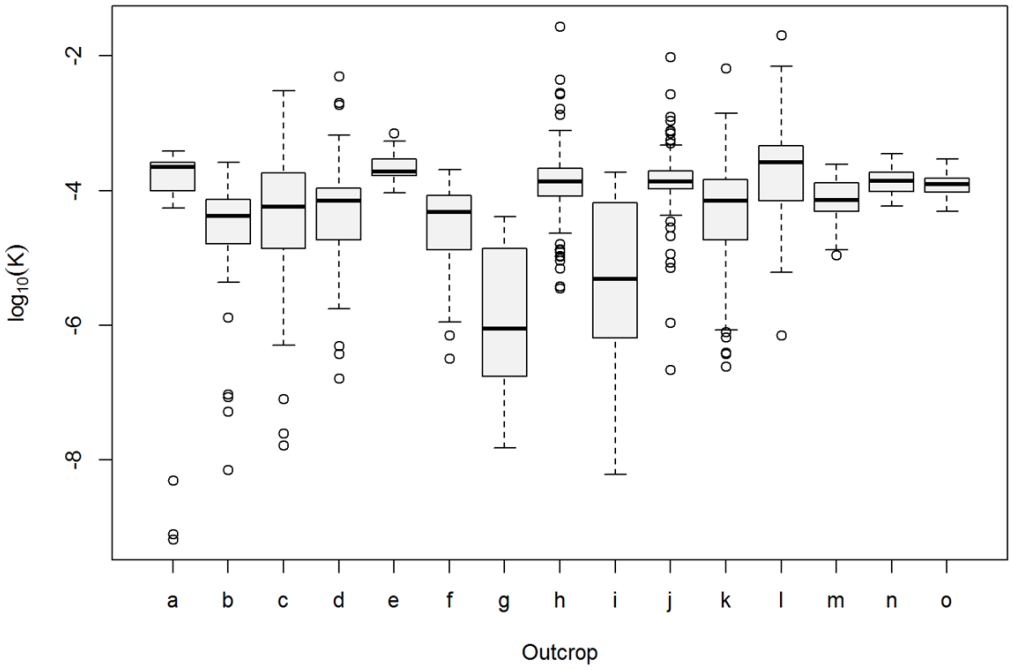

Figure 3.5. Boxplots of the air permeability-derived *K* values for 15 outcrops (see Table 3.2 for outcrop codes a-o).

By repeating measurements on multiple samples, with $\log_{10}K$ between -3.5 and -6.5, the measurement error was determined to correspond to a geometric CV of about 2% (absolute value taken). The variability observed within each outcrop exceeds the variability owing to measurement error (Table 3.2). This indicates that the observed variability is mainly due to the sediment characteristics. For the second Bolderberg Sands outcrop (o), the measurement error may be up to half of the observed variability. Overall the measurement error is very small; hence it is not further treated separately.

Since the measurements of most outcrops were performed on regular grids with knowledge of the relative 2D coordinates, experimental variograms were calculated to describe the spatial structure within these outcrops. Outcrop a was omitted as no regular grid was investigated there, and outcrop b and i are not included in the horizontal variography due to the limited horizontal extent of the investigations. Directional experimental variograms for the horizontal (parallel to stratifications) and vertical (orthogonal to stratifications) directions are presented in Fig. 3.6.

3.3.1. Quaternary sediments

The Quaternary stratigraphy of Belgium is essentially made up of continental facies deposited by fluvial, eolian, slope or organo-chemical processes (Gullentops *et al.*, 2001). The sediments studied in this paper are estuarine to fluvial, exhibit fine-grained clayey to sandy textures, and belong to the Weelde or Sterksem Formations, or represent alluvial deposits by the Kleine Nete river.

3.3.1.1. The Schelde Group

The Schelde Group includes all fluvial deposits in the Schelde and Yser basins. The oldest one corresponds to terraces formed during valley incision in Lower and Middle Pleistocene. Younger deposits are present in the valley bottom fill. Deposits occurring as far upstream as the Kleine Nete basin probably originate from Late Glacial to Holocene filling of Late Glacial incisions (Beerten, 2010).

Outcrop a: Kleine Nete deposits at Kasterlee

Paleo-channels of the Kleine Nete river deposits together with fluvial point bar deposits were investigated near the city of Kasterlee. Fluvial point bar sands (Fig. 3.7A,B), a peat paleobedding (Fig. 3.7C) and a more recent

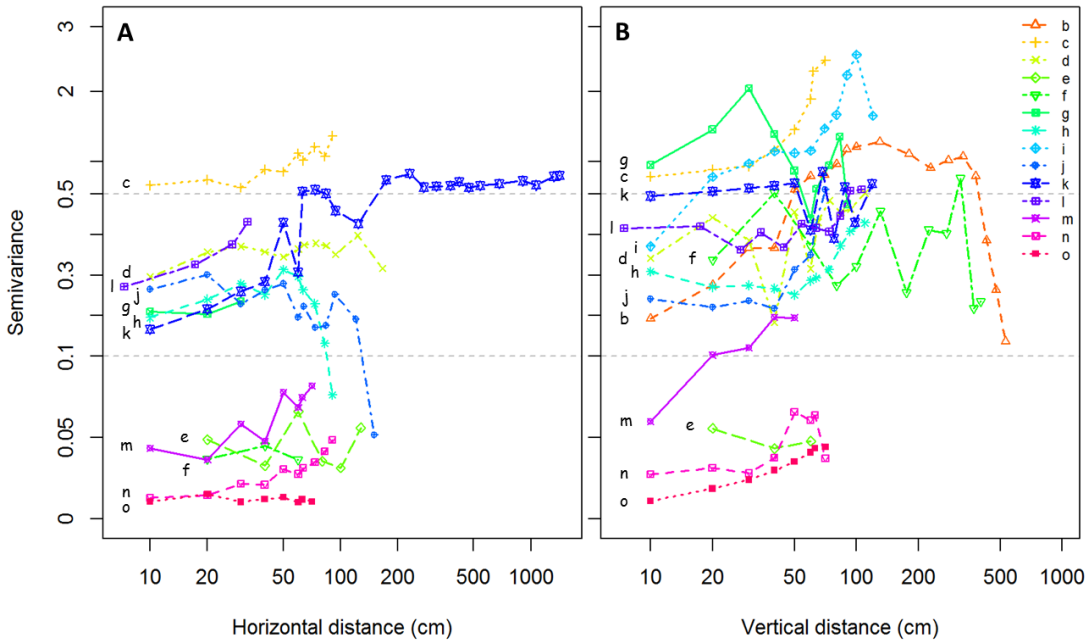

Figure 3.6. Directional (A: horizontal; B: vertical) experimental variograms for the outcrops (see Table 3.2 for outcrop codes a-o). Note the y-axis is a stack of three different scales.

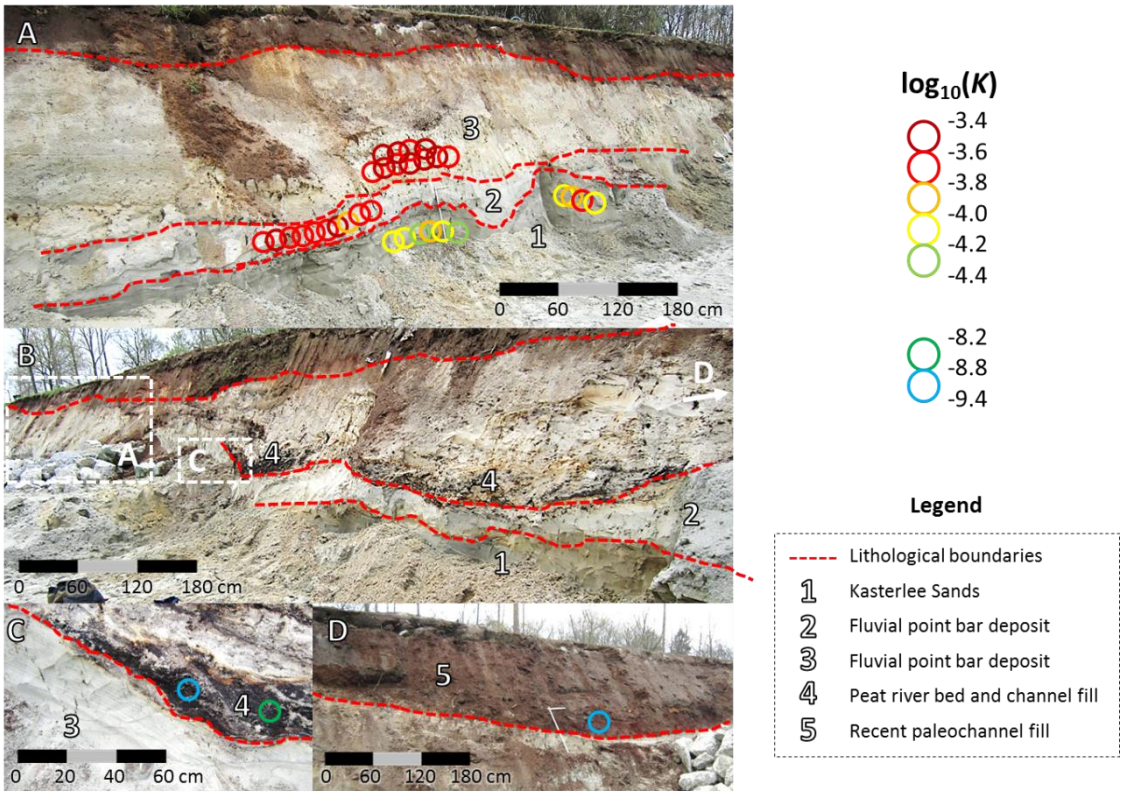

Figure 3.7. Kleine Nete paleochannel outcrop, with several fluvial point bar sand deposits (A, B), an old paleochannel (C), and the more recent remains (1950s-1970s) of a backfilled meander (D). Positions of the measurements are approximate, as quantification of the contrast between the different units was the main target here.

channel fill (Fig. 3.7D) that finds its origin in the straightening of the Kleine Nete river from the 1950s to the 1970s (De Bie *et al.*, 2007; AMINAL, 2003) are clearly distinguishable. Grain size analysis (Fig. 3.8) shows point bar sand deposits have more than 95% sand and less than 1% clay. The percentage of fines increases to a few percentages in the lowermost sand unit, which corresponds probably to the Tertiary Kasterlee sands. The peaty river bed and the red-coloured material in the recent channel fill (Fig. 3.7C, D, and uppermost two samples in Fig. 3.8) show a few percentages of clay and almost 60% of silt.

Air permeability-derived *K* values agree reasonably well with predictions based on particle size data; details about the multiple linear regression predictive model (MLR) and the generalized likelihood uncertainty estimation - artificial neural network model (GLUE-ANN) are available in Rogiers *et al.* (2012; Chapter 6). The largest difference between measured and predicted *K* is for the top of the river bed sediment (Fig. 3.8). A high moisture content in the river bed samples likely influenced air permeability measurements resulting in low *K* values. For the point bar sands, this offset seems to be no more than approximately one order of magnitude. Such discrepancies were also reported by Rogiers *et al.* (2013b; Chapter 4) with regards to a comparison between outcrop and subsurface values; the main reason being that the grain size based *K*-prediction model was developed using subsurface sediments exhibiting slightly different grain size properties than the outcrops (Chapter 6). Gridded measurements were not performed on this outcrop, hence no spatial analysis was carried out.

The point bar sands proved to be very homogeneous, as all *K* values from the air permeameter measurements are within one order of magnitude (Fig. 3.8). A mean *K* value of $10^{-3.7}$ m/s indicates very permeable sands (Table 3.2), hence current surface water bodies that are eroding such sands might show a high degree of connectivity to the groundwater. In contrast, both the channel fill and peat bedding display relatively low *K* values ($10^{-8.9}$ m/s; such low values are influenced by the high moisture content (saturation degree > 80%) in these materials; based on a comparison with independent *K* measurements, the grain size-based predictions are more accurate; see also Rogiers *et al.* 2013a; Chapter 2). Regions where these deposits exist in the river bed are likely to have a low degree connectivity with the groundwater. As the measurements at this outcrop were exploratory, and not performed on a regular grid, this outcrop is not included in the upscaling approach.

3.3.1.2. The Weelde Formation

The Weelde Formation is part of the Campine Complex (also known as the Kempen Group), and consists of fine grained sandy to clayey sediments, all of which are the result of alternating estuarine and fluvial-eolian conditions at the moment of deposition (Gullentops *et al.*, 2001).

Outcrop b: Weelde Formation at Beerse

The Weelde Formation is accessible through several brickworks quarries at the interfluves between the Nete and Meuse basins in the northern part of the study area. The top 5.5 m of the formation was studied at the Beerse clay pit because the lithological succession could serve as analogue for engineered earth covers whose design is being optimized (Jacques *et al.* 2010). The 5.5-m deep vertical profile starts at the bottom in the top of a clay layer and is cross-cutting a second clay layer, as shown in Fig. 3.9. Grain size analyses for this profile (Fig. 3.10) reveal the clay percentage ranges from 6 to 18% in the clayey layers, with a pronounced silt content for the central layer (up to 65%). A thin layer of coarse sands (at 40 and 180 cm) is present on top of each of the clay layers, which creates a large difference in lithology on vertical distances of only a few centimetres.

The air permeability-based *K* values (Fig. 3.10) clearly show the presence of the central clay layer between approximately 1.6 and 2.6 m. Only a few measurements were performed in the clay itself because of the long equilibration time required to obtain a reliable value; the minimum value measured was $10^{-8.1}$ m/s (Table 3.2). This low value was used to fill in missing values at several measurement depths within the clay layer on the 10 cm-spaced profile thus facilitating the upscaling later on (Fig. 3.4b). The grain size based *K* predictions show a similar pattern, but with an offset of about one to two orders of magnitude towards lower values. As indicated above, this offset corresponds to the outcrop – subsurface sediments bias observed by Rogiers *et al.* (2013b; Chapter 4), probably caused by

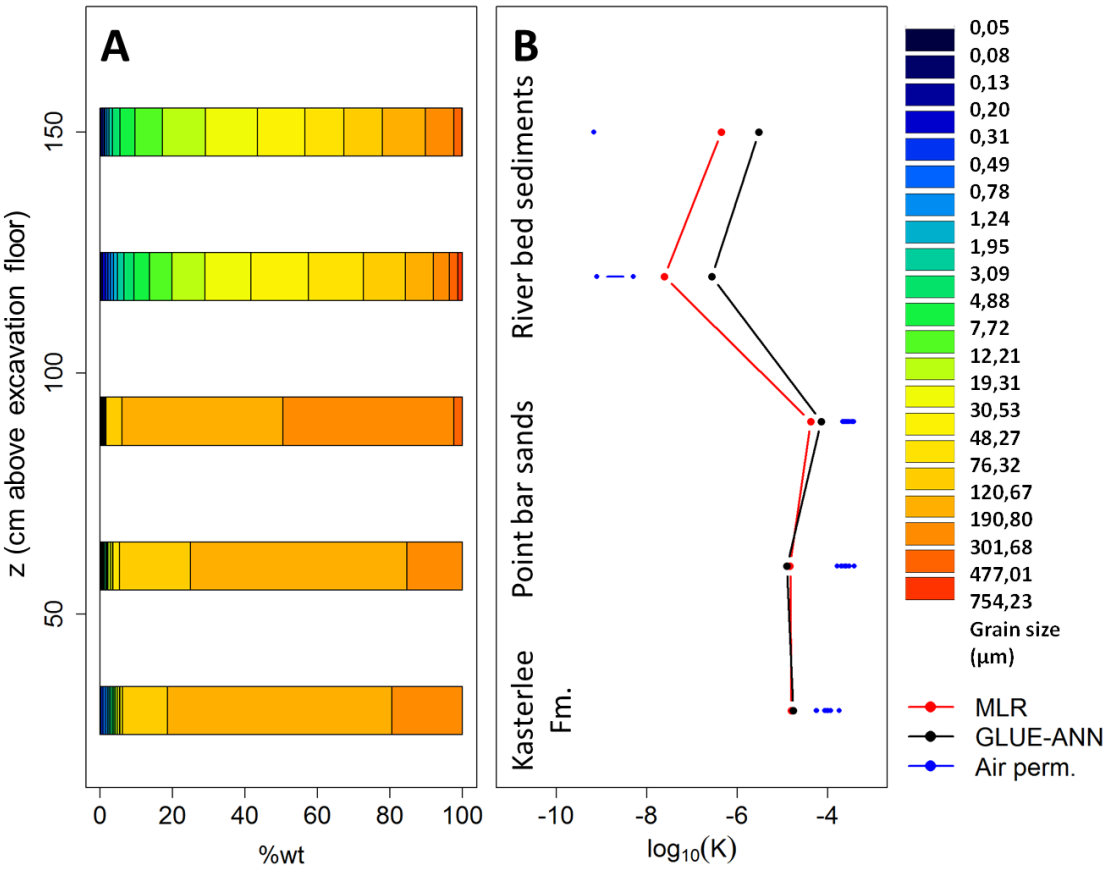

Figure 3.8. Grain size distributions (A) and hydraulic conductivities (B) derived from the air permeameter measurements (Air perm.) on the same units, and predicted with the multiple linear regression model (MLR) and GLUE-ANN ensemble of Rogiers *et al.* (2012; Chapter 6) for the Kleine Nete deposits outcrop. The vertical position is approximate, as not all samples were taken at the same position. The soil surface is approximately 2.5 m above the excavation floor.

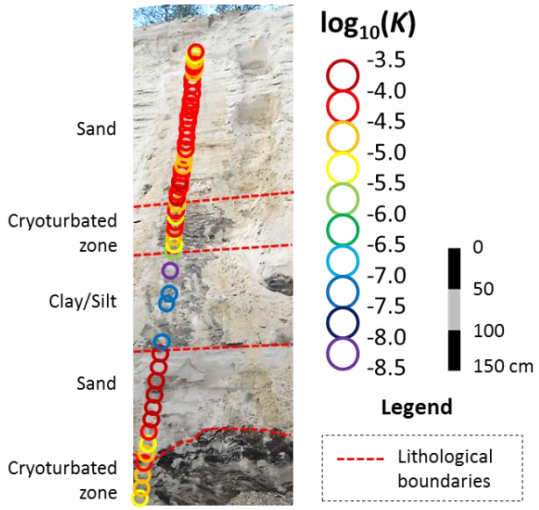

Figure 3.9. Campine Complex outcrop.

mineralogical weathering processes, physical degradation at the macroscopic scale, differential compaction histories, all of which are not identifiable on the basis of grain size. The grain size based predictions are not consistent for the samples with a very large silt fraction, since these were not encountered in the borehole data from which the models were derived (see Rogiers *et al.*, 2012; Chapter 6). It is worth noting that the extrapolation to these silty grain size distributions by the linear MLR model is more trustworthy than the non-linear extrapolation of the GLUE-ANN model. In terms of spatial correlation structure, the vertical experimental variogram shows clearly the influence of the presence of a clay layer in the central part of the profile (Fig. 3.6, outcrop b). The small-range variability is comparable to that of the other sandy outcrops studied. The entire vertical profile is used to calculate outcrop-scale parameter values in section 3.4.

Outcrop c: Cryoturbation and ice wedges in the Weelde Formation, Beerse

A more detailed investigation of the bottom layers of the Campine Complex outcrop (lowest cryoturbated zone in Fig. 3.10) identified cryoturbation and ice wedges that have disturbed the otherwise sharply delineated clay layer (Fig. 3.11). The study of such features is relevant since it will alter the hydraulic properties of the clay layer, and quantification of such impacts provides useful underpinning of studies of long-term performance of engineered clay barriers.

This part of the outcrop with its characteristic frost structures exhibits a larger variability in K than the measurements in the sediment succession above (Fig. 3.5, outcrop c versus b). The past mixing of sediments owing to these paleo frost processes will have contributed to this variability with a data range of over five orders of magnitude (Table 3.2), both in horizontal and vertical direction. Due to cryoturbation and formation of ice wedges and their infillings the spatial rearrangement of the original stratification has been disturbed; as a result this outcrop shows the highest semivariance in horizontal direction for all lag distances (Fig. 3.6, outcrop c). Several measurements in the clay are missing due to excessively long equilibration times. Therefore the minimum observed value is used to complete the 10-cm spaced grid for the purpose of the upscaling calculations in section 3.4 (shown in Fig. 3.4, outcrop c).

3.3.1.3. The Sterksel Formation

Coarse gravelly sands were deposited by the Rhine as a braided river (Beerten, 2006), attracted into the Roer Valley Graben by its renewed sinking (Gullentops *et al.*, 2001). The Sterksel Formation occurs in the lower Meuse valley, and is of Early Pleistocene age. The Lommel Member, that covers the Western part of the Kempen, was attracted by the antithetic down warping along the Rauw Fault.

Outcrop d: Lommel Sands at Lommel Maatheide

The Lommel Sands were studied at the Maatheide quarry of Sibelco. They are part of the unsaturated zone through which groundwater recharge takes place; the water table is about 20 cm below the bottom of the outcrop. The thickness of Lommel Sands is up to several meters in the east of the study area (Gullentops & Wouters, 1996). Cross-lamination stratification and fluviatile gully structures are clearly visible on Fig. 3.12. Within the underlying Mol Sands, a compact lignite layer is present that represents an important hydraulic barrier.

Air permeability measurements were performed with a spacing of 10 cm on a 2 by 1.8 m grid. This resolution might be insufficient to investigate the fine-scale layering within the cross-bedded fluviatile gullies (about half a meter high and a few meters wide), but a clear contrast exists at the boundaries between gullies, which is also indicated by the visual colour contrasts and the air permeameter measurements (Fig. 3.12). Hydraulic conductivity clearly varies over several orders of magnitude within the Lommel Sands, even within one single gully (Table 3.2; Fig. 3.5). There are missing K values at two depths: the sandy material at the middle part of the outcrop was too coarse and loose for making reliable measurements; furthermore, the bottom part was too wet owing to the shallow groundwater table. Three measurements could however be performed within the lignite layer, with an average of $10^{-5.8}$ m/s (not shown). This indicates that these 0.5- to 1-m thick compacted banks, with large spatial continuity, only disturbed by faults, function as important hydraulic barriers within the Mol Sands.

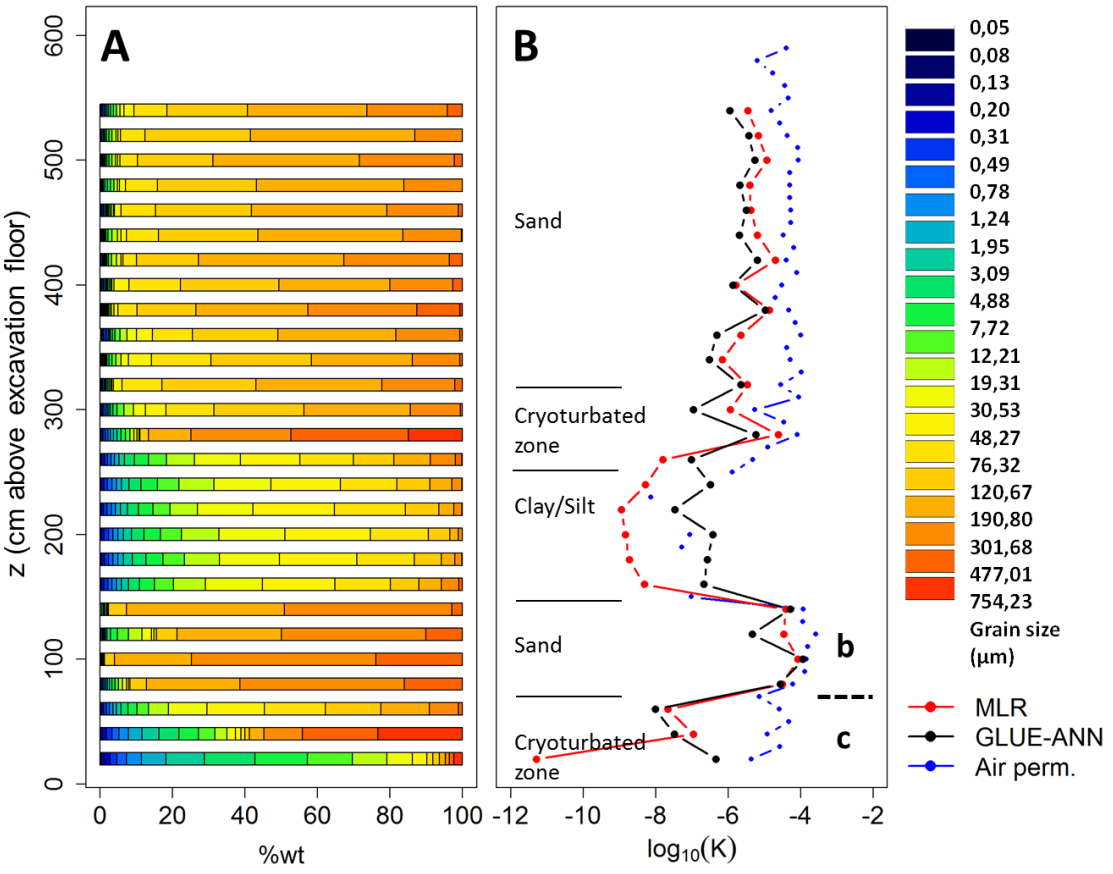

Figure 3.10. Grain size distributions (A) and hydraulic conductivities (B) derived from the air permeameter measurements (Air perm.), and predicted with the multiple linear regression model (MLR) and GLUE-ANN ensemble of Rogiers *et al.* (2012; Chapter 6) for the vertical profiles through the Campine Complex outcrop.

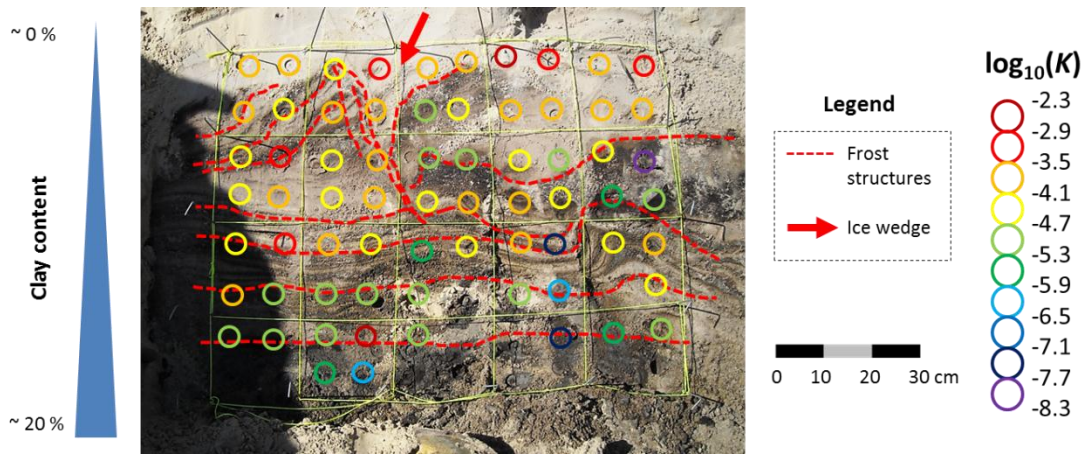

Figure 3.11. Cryoturbation and ice wedge in the top of a clay layer in the Weelde Formation outcrop. This clay layer appears at the bottom of the profile shown in Fig. 3.9 with sample locations shown as open circles.

The spatial variability seems to be more or less equal in horizontal and vertical directions, and there is little spatial structure over the short distances investigated (Fig. 3.6, outcrop d). Further upscaling analyses are done in section 3.4 on the upper 60 cm of the measurement grid within the Lommel Sands (Fig. 3.4d), because several measurements are missing at depths greater than 60 cm below the top of the grid, and the lignite is considered to be a quite homogeneous unit. One single missing value was interpolated from its neighbours to complete the grid.

3.3.2. Pliocene sediments

3.3.2.1. The Mol Formation

The Tertiary Mol Formation consists of white, pure (99% SiO_2) (Sibelco, 2010), coarse and medium fine continental/estuarine sands, sometimes lignitic and with some lenses of micaceous clay (Laga *et al.*, 2001). The bottom part is very slightly glauconiferous (less than 2 %) (Beerten *et al.*, 2010). Within the geographic boundaries of the Mol municipality, the 20–m thick Mol Formation is divided into the Upper Mol and Lower Mol Sands, both Pliocene in age (Beerten *et al.*, 2010). The latter are very well sorted and finer and darker in colour, while the former are moderately to well sorted medium sands with a basal gravel layer. Because of the high siliceous content of this sand, it is being mined for various industrial uses (Gullentops and Wouters, 1996). One outcrop site was investigated at the municipality of Lommel.

Outcrop e: Mol Sands at Lommel

These Pliocene sands occur one to two meters below the water level of the dredging pond at the Maatheide quarry, but the depth of the Quaternary-Tertiary boundary shows considerable variation. A clear boundary was observed just below the ground surface, and a small outcrop was excavated, as shown in Fig. 3.13. Except for some small oxidized horizontal bands, no sedimentary structures could be distinguished in these Mol Sands.

Because the material is very loosely packed, a sampling density of 20 cm spacing was applied to prevent collapse of the excavation. The average K value amounts to $10^{-3.7}$ m/s, which indicates a very permeable material (Table 3.2). Spatial correlation is hard to detect (Fig. 3.6, outcrop e), but the variability within this outcrop clearly is very small with all K values within a range of less than one order of magnitude The difference between these K values and those of the underlying lignite layer (see Lommel Sands outcrop) again underlines the role as hydraulic barrier of the latter ($10^{-3.7}$ versus $10^{-5.8}$ m/s). These continental/estuarine sands seem to compare very well in hydraulic characteristics sense with the Kleine Nete point bar deposits, with a very similar CV (Table 3.2, Fig. 3.5, outcrop a and e) and only slightly visible sedimentary structures (*e.g.* point bars left of the peat in Fig. 3.7C). The difference with the fluviatile Lommel Sands ($10^{-3.7}$ for Mol Sands versus $10^{-4.21}$ m/s for Lommel Sands) is probably due to the different grain size characteristics (Lommel being more heterogeneous and less well sorted than Mol) and depositional conditions (fluviatile versus estuarine). The entire measurement grid is used for upscaling calculations in section 3.4, as shown in Fig. 3.4.

3.3.2.2. The Poederlee Formation

The Poederlee Sands are fine, slightly glauconiferous sands, with small lenses of clay in the lower part (Laga *et al.*, 2001). They are lateral equivalents of the Mol Sands (Vandenberghe *et al.*, 2004; Louwye & De Schepper, 2010). The oldest depositional environment of the Poederlee Formation was neritic with an open ocean influence, later changing into more shallow waters in the proximity of the coast with fresh water influence (Louwye & De Schepper, 2010). Deposition likely took place during a transgressive to highstand phase, under a high and possibly maximal sea level, until the accommodation space was filled. The Poederlee Sands were studied at two locations, *i.e.* Kasterlee (Figs 3.14-16) and Lichtaart (Fig. 3.17).

Poederlee Sands at Kasterlee were analysed in two parts: an upper part (denoted further on as outcrop f) includes two sampling zones (Figs 3.14A, B & C) and the lower part (denoted further on as outcrop g) encompasses one sampling zone (Figs 3.14A & 3.15).

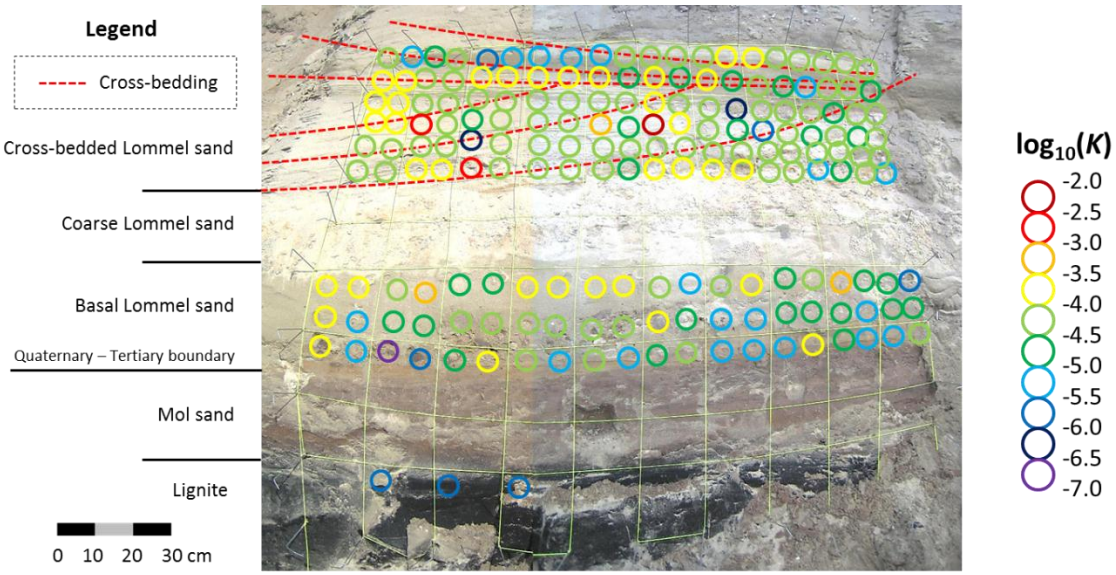

Figure 3.12. Lommel Sands outcrop showing the Quaternary-Tertiary boundary; also visible are the top of the Mol Formation represented by layers of Mol sand and lignite. The upper six rows of measurements were used for the calculations.

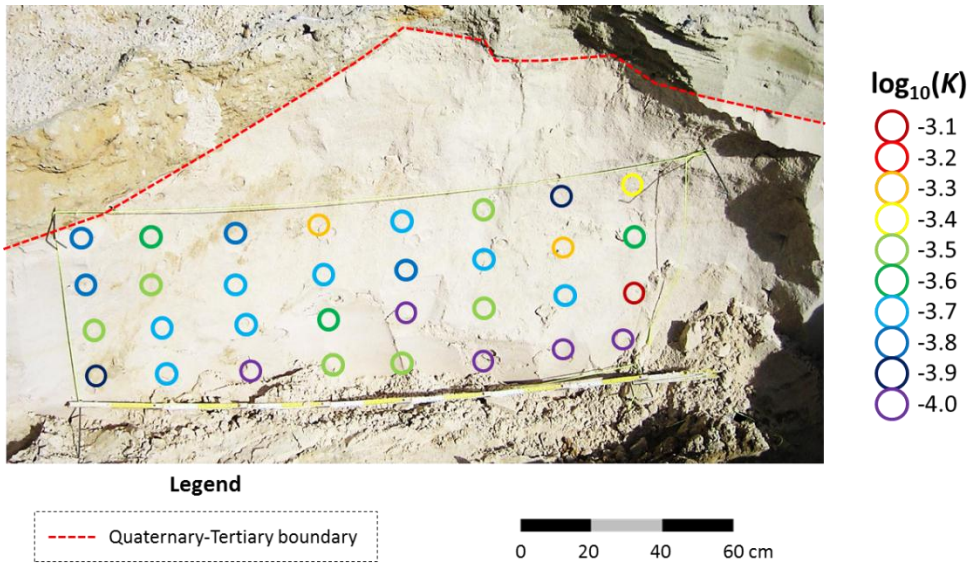

Figure 3.13. Mol Sands outcrop showing the Quaternary-Tertiary boundary.

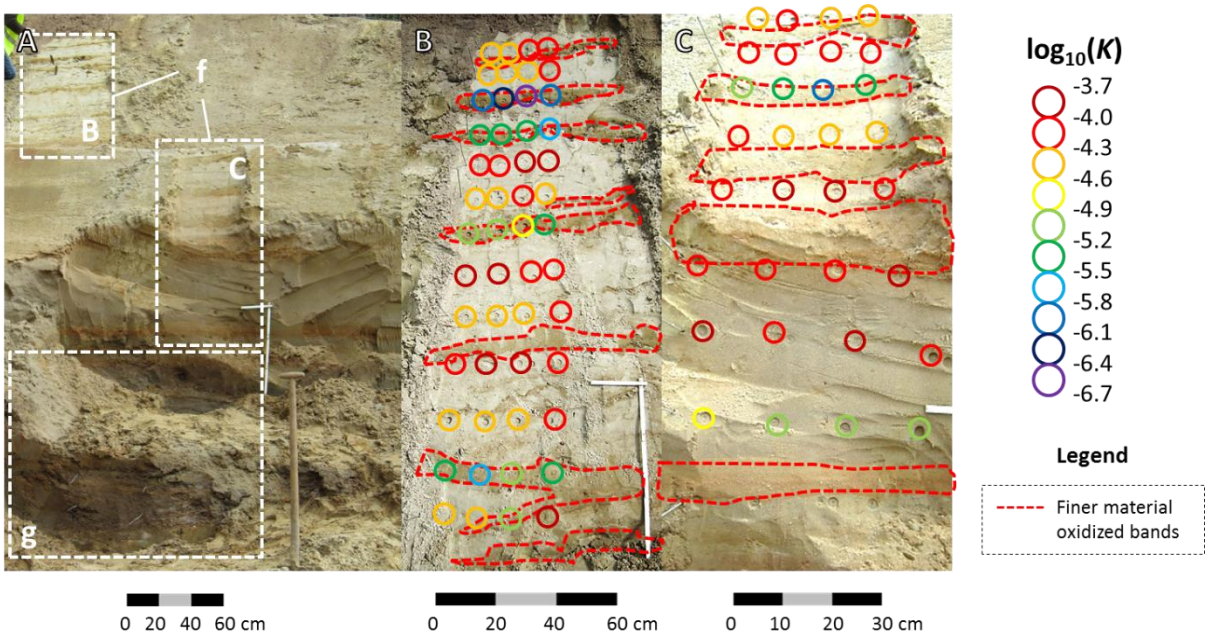

Figure 3.14. Overview (A) of three sampling zones; top part of the Poederlee Sands outcrop at Kasterlee with distinct fine-grained oxidized bands (B, C), denoted as outcrop f.

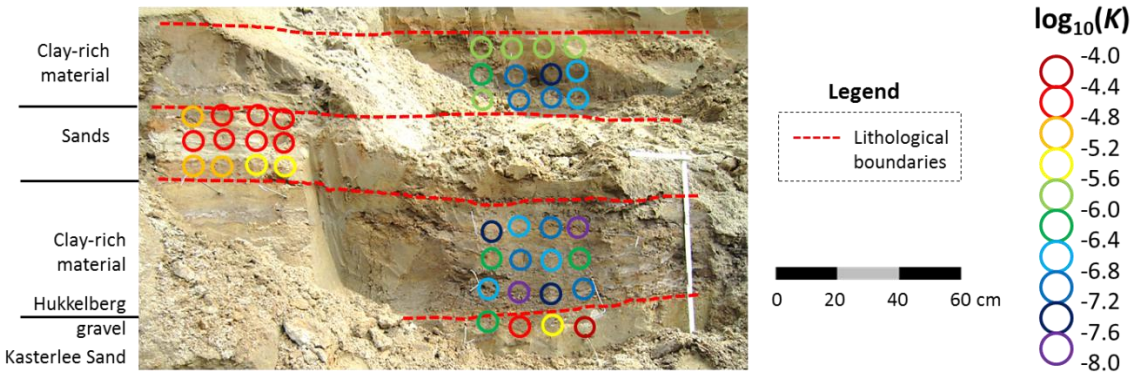

Figure 3.15. Bottom of the Poederlee Formation at Kasterlee (outcrop g) with several pronounced clay lenses; the Hukkelberg basal gravel indicates the boundary with the Kasterlee Sands.

Outcrop f: Poederlee Sands at Kasterlee (top of outcrop)

This upper part of the Poederlee Formation exhibits several bands of fine oxidized material within a background of white fine sand, with sand contents of respectively 70 and 90% (Figs 3.14B, C & 3.16).

This part of the Poederlee Sands at Kasterlee shows a mean K value of $10^{-4.5}$ m/s with a standard deviation of 0.61 (Table 3.2). The measurements were done with 20 cm spacing in vertical direction and 10 to 15 cm in horizontal direction. The variability expressed in terms of a three orders of magnitude range is caused by the presence of the horizontal bands composed of finer-textured material within the more permeable sand; lower K values typical of the finer-textured bands are clearly visible in Figs 3.4 (outcrop f) and 3.14. As a result the horizontal spatial variability is considerably smaller than the vertical one (sill for horizontal variogram is ~0.05 versus ~0.4 for the vertical) (Fig. 3.6). All measurements in Fig. 3.14B & C are combined (see Fig. 3.4) and used to calculate outcrop-scale parameter values in section 3.4.

Outcrop g: Poederlee Sands at Kasterlee (bottom of outcrop)

The bottom part of the Poederlee Sands outcrop at Kasterlee (Figs 3.14 & 3.15) exhibited layers with increased clay and silt content (resp. 6 and 30%), ranging in thickness from a few centimetres to several tens of centimetres. Presence of the basal Hukkelberg gravel (Laga *et al.*, 2001) indicated that these sediments correspond to the bottom of the Poederlee Formation.

The average *K* value for outcrop g is $10^{-5.94}$ m/s, and the standard deviation (1.07) is one of the highest observed (Table 3.2). Also worth noting is that the maximum value is the lowest maximum observed across all outcrops (Fig. 3.5).

The few hydraulic conductivities predicted from the grain size analyses of outcrops f and g confirm the air permeameter-derived values (Fig. 3.16). Although only three grain size predicted *K* values exist, it is useful to discuss the absence of an offset which was observed for other outcrops. A reason for this might be the fact that the outcrop was only very recently excavated for a tunnel project, and there are still a few meters of overburden above the highest measurement point. As a consequence, the sediments have not been influenced by weathering processes, physical degradation of sediment structure, migration of clay particles, root penetration, etc. For a comparison between outcrop-derived and hard subsurface data, the reader is referred to Rogiers *et al.* (2013b; Chapter 4). At Kasterlee, the basal part of the Poederlee Formation thus represents potentially a very important hydraulic barrier. The horizontal spatial variability seems to be moderate (sill ~0.2), while the vertical variability seems to be very high at short ranges (sill ~1-2) (Fig. 3.6). Moreover, the very large excavation in the framework of a tunneling project allowed to follow the connected network of clay lenses at least along 150 m. As the main heterogeneity occurs in the vertical direction, all measurements are combined into a 10-cm spaced grid, as shown in Fig. 3.4, for the upscaling calculations in section 3.4.

Outcrop h: Poederlee Sands at Lichtaart

At the Poederlee Sands outcrop in Lichtaart we observed oxidized horizons several cm-thick in addition to extensive bioturbation features and the Hukkelberg base gravel (Fig. 3.17). Sedimentary structures such as small-scale foresets and bottomsets were not distinguishable. The clay content increases towards the bottom of the formation, in the form of a few distinct small clay lenses.

While the Poederlee Sands outcrops at Kasterlee (Fig. 3.15) and Lichtaart (Fig. 3.17) are about 2.5 km apart, there is a noticeable difference in the geometry of the clay lenses in the bottom of the Poederlee Formation. In Kasterlee the clay lenses are fewer in number, up to tens of centimetres thick and appear to be connected, while in Lichtaart there are many disconnected lenses of less than a centimetre thick.

A clear vertical trend is visible in the *K* matrix of the outcrop (see Fig. 3.4), which was measured with a 10 cm spacing. This reflects the increasing clay content and heterogeneity towards the base of the Poederlee Sands at this site. The volume occupied by clay lenses is however very different from that at the bottom of the formation at Kasterlee. A significant difference in *K* for the brown-coloured oxidized horizons is not present, although at some levels the conductivity is slightly lower in such horizons. The overall mean K value is 10-3.9 m/s, with a relatively small log10(K) variance of 0.28 (Table 3.2). The log10(K) range however goes up to 3.88 due to the trend in the bottom of the outcrop. The horizontal direction exhibits spatial structure and the vertical short-range variability is rather high (Fig. 3.6, outcrop h). The entire measurement grid is used to calculate outcrop-scale parameter values in section 3.4.

Outcrop i: Poederlee-Kasterlee transition zone, Lichtaart

The Hukkelberg gravel is the compact basal gravel bed of the Poederlee Sands and forms the transition towards the underlying Kasterlee Sands (Fig. 3.18). It consists of rounded quartz and flint pebbles and silicified carbonates (Laga *et al.*, 2001). Above this layer, the clay content again clearly shows elevated values in the bottom of the Poederlee Formation (Fig. 3.19). The gravel layer itself is compacted and cemented and includes several concretions up to several centimetres in size. Such a layer increases the anisotropic nature of the hydraulic conductivity tensor and thus has implications for the water flow across this stratigraphical boundary. Currently the concretions occur in the unsaturated zone; however several borehole cores from aquifer sediments (Beerten *et al.*, 2010) also contained similar concretions

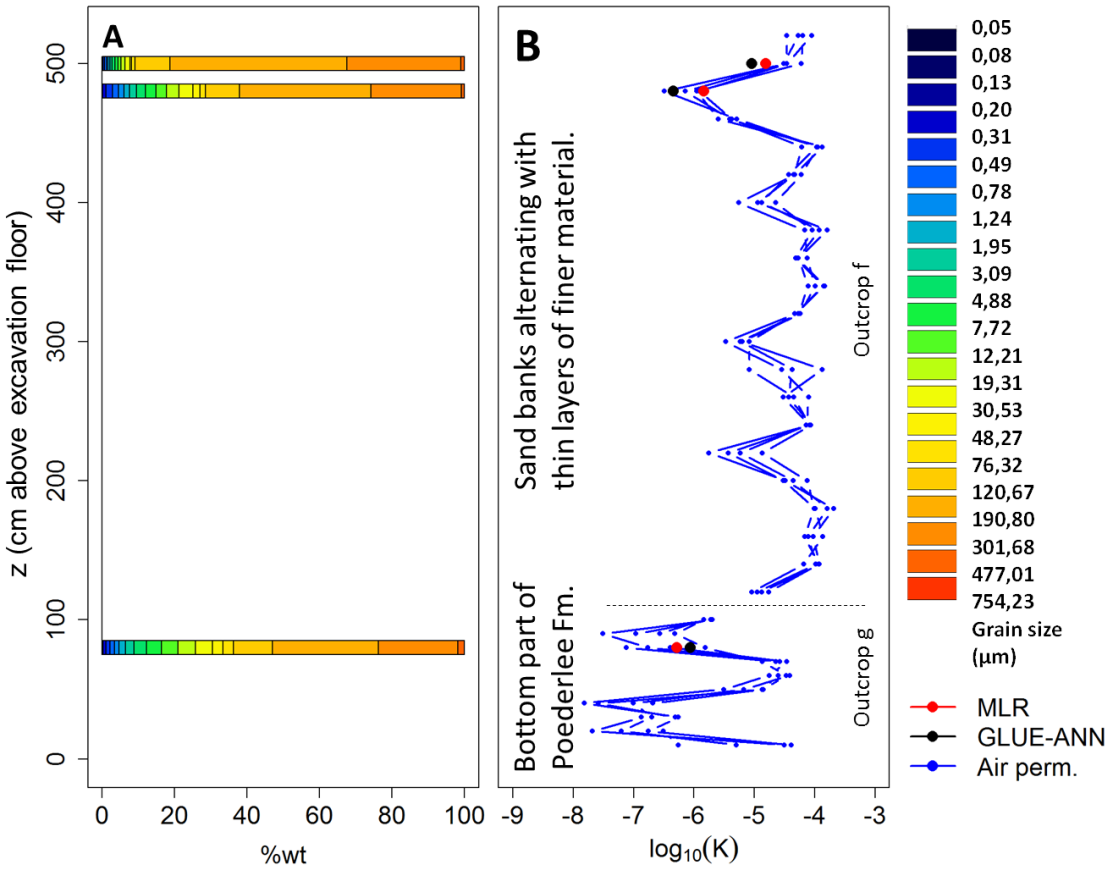

Figure 3.16. Grain size distributions (A) and hydraulic conductivities (B) derived from the air permeameter measurements (Air perm.), and predicted with the multiple linear regression model (MLR) and GLUE-ANN ensemble of Rogiers *et al.* (2012 (Chapter 6)) for the Poederlee outcrop at Kasterlee. The connected dots represent vertical transects.

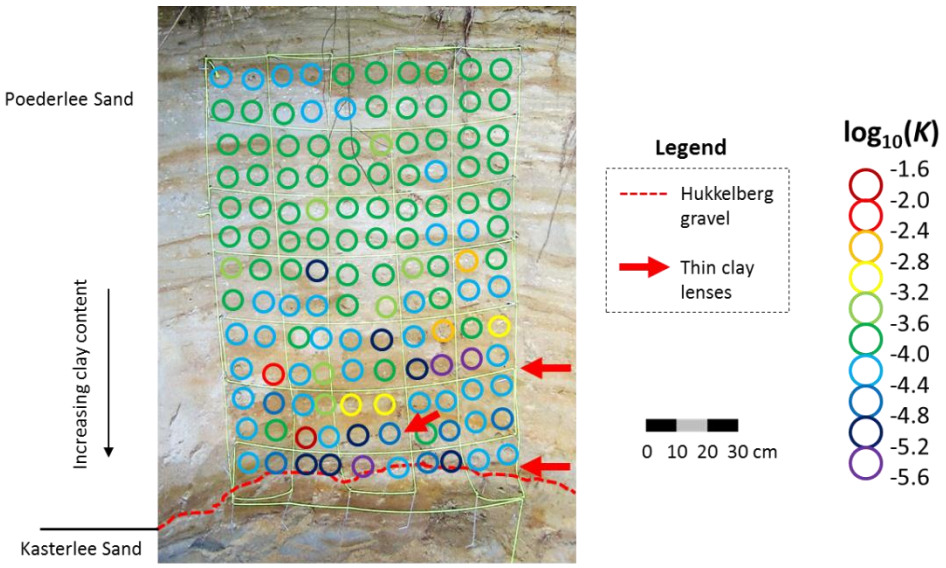

Figure 3.17. Bottom of the Poederlee Formation at Lichtaart, including a few very thin clay lenses (see arrows), and the Hukkelberg basal gravel.

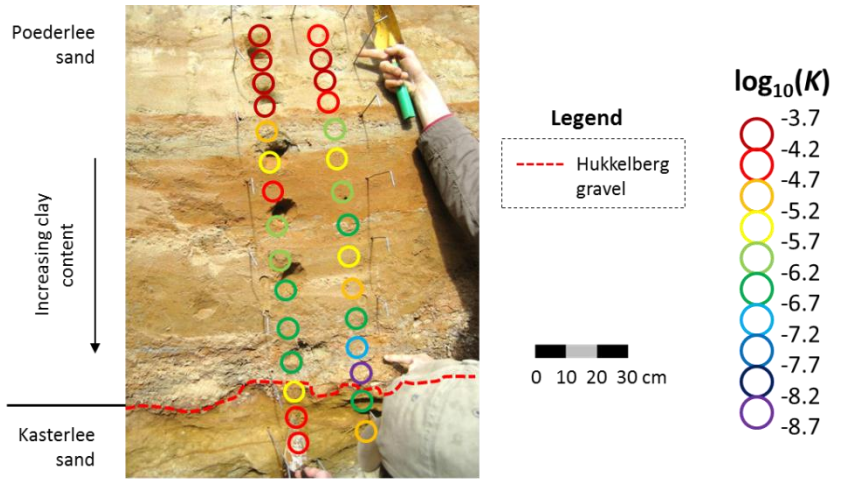

Figure 3.18. Transition between the Poederlee and Kasterlee Sands at Lichtaart including the Hukkelberg gravel stratigraphic boundary.

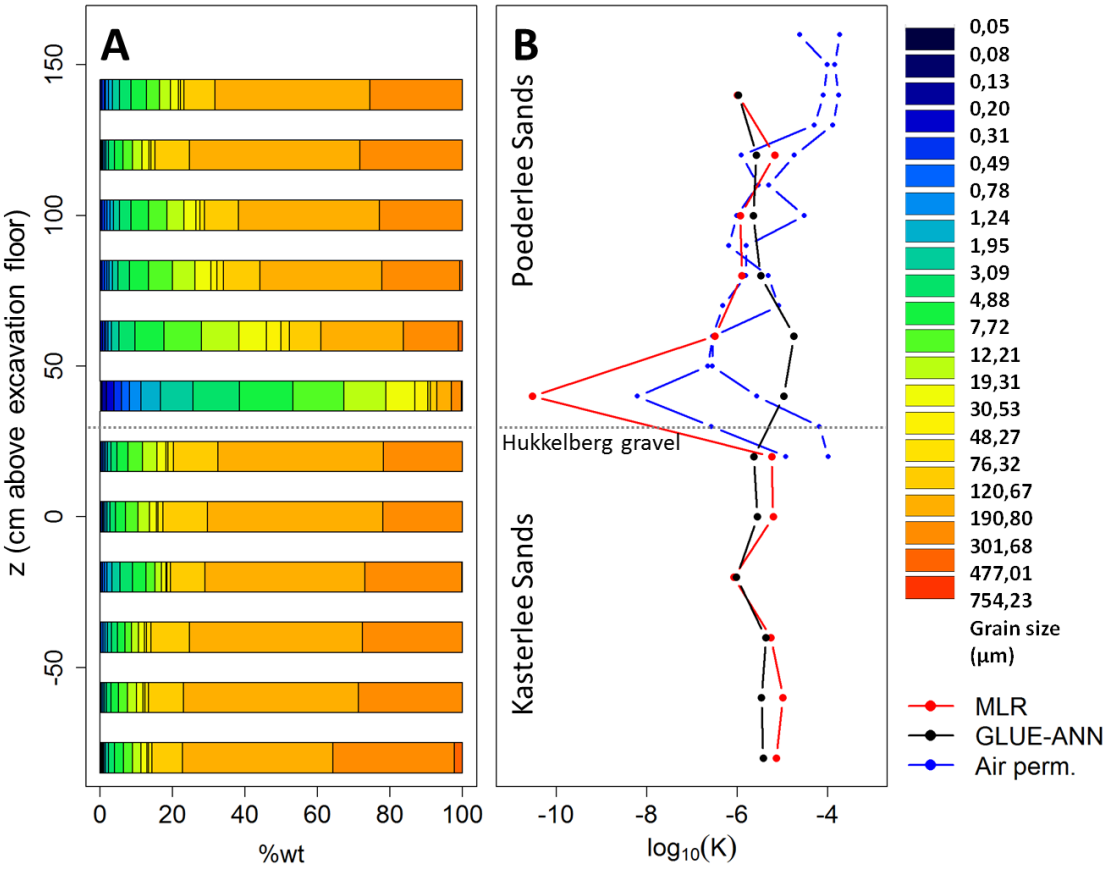

Figure 3.19. Grain size distributions (A) and hydraulic conductivities (B) derived from the air permeameter measurements (Air perm.), and predicted with the multiple linear regression model (MLR) and GLUE-ANN ensemble of Rogiers *et al.* (2012; Chapter 6) for the Poederlee-Kasterlee transition.

indicating such structures might have formed at different times during the Poederlee-Kasterlee and Kasterlee-Diest hiati in palaeo-unsaturated conditions.

The air permeameter measurements were performed across the stratigraphical boundary on two profiles 30 cm apart using a vertical spacing of 10 cm. The bottom of the Poederlee sands shows an increasing clay content towards the compacted and cemented gravel layer and hence displays lower *K* values (Figs 3.4, 3.18 & 3.19). This outcrop has a mean *K* value of $10^{-5.28}$ m/s and the highest variance of *K* (1.29 see Table 3.2), which can be attributed to grain size variations (lower *K* due to increasing clay content), post-depositional oxidation (lower *K* with increasing oxidation features) and cementation processes yielding lower *K*. The Hukkelberg gravel bed is very thin and not always continuous. Hence it is not likely to influence the hydraulic properties across this profile. The vertical variability increases with distance until about 1 m (Fig. 3.6B). All measurements displayed in Fig. 3.18 are used to calculate outcrop-scale parameter values in section 3.4.

The grain size analyses of this section of the profile are presented in Fig. 3.19. The samples are taken more or less each 20 cm. The increasing amount of silt and clay content towards the base of the Poederlee Sands is very clear. Below the base, the sand content increases to more than 80% in the Kasterlee unit. The Hukkelberg gravel itself was not sampled since it was very thin, and probably contains too much coarse fragments for grain size analysis using the laser diffractometer. Note that all grain size distributions are bimodal in this profile, with virtually no particles between 30 and 50 µm. The grain size based hydraulic conductivity predictions correspond reasonably well with the air permeameter based values. Only the sample at the base of the Poederlee Formation, exhibiting a very high fine silt content, seems to have lower conductivities based on the MLR grain size model. This clearly shows that physical degradation of the sediment structure has a larger effect on the clayey parts of the studied outcrops than the fraction of sand particles. The GLUE-ANN model fails to make correct predictions since such high fine silt fractions were not present in the borehole data used to generate (*i.e.* train) the model (Rogiers *et al.*, 2012; Chapter 6).

3.3.3. Miocene sediments

3.3.3.1. The Kasterlee Formation

The Kasterlee Formation consists of a relatively homogeneous fine, micaceous, slightly glauconitic, sandy upper part (Laga *et al.*, 2001), and a very heterogeneous alternation of clay lenses and sand banks in the lower part (Beerten *et al.*, 2010). The homogeneous upper part is referred to as Kasterlee Sands while the heterogeneous lower part is named Kasterlee Clay. The depositional environment for this formation is coastal or a coastal embayment, with manifest river influence (Louwye *et al.*, 2007).

Outcrop j: Kasterlee Sands at Lichtaart

Based on analysis of hand-augered boreholes several meters deep, the site at Lichtaart shows no presence of the clay lenses within the bottom part of the Kasterlee Formation. This outcrop (Fig. 3.20) can therefore only be used as an analogue for the Kasterlee Sands. Except for some oxidized horizons, sedimentary structures were difficult to distinguish, suggesting a fairly homogeneous material commensurate with the typical lithological characteristics of Kasterlee Sands.

The Kasterlee Sands show a mean *K* value of $10^{-3.9}$ m/s at the Lichtaart site (Table 3.2). While the standard deviation is limited (0.5), several outliers are present (Fig. 3.5), causing a *K* range of over four orders of magnitude. The lack of clear spatial structures or trends is also visible from the pure nuggets in the experimental variograms (Fig. 3.6). The entire measurement grid is used to calculate outcrop-scale parameter values in section 3.4.

Outcrop k: Kasterlee Clay at Heist-op-den-berg

The exposures on the Heist-op-den-berg and Beerzelberg hilltops are useful analogues for the Kasterlee Clay as previous studies identified the presence of the Kasterlee Formation with several thick clay lenses (Fobe, 1995). We studied the sediments in a sunken lane towards the hilltop providing exposure over several tens of meters. Clay lenses were found to be continuous at the scale of the outcrop (about 16 m, Fig. 3.21A & B). The lenses were heavily affected by physical weathering (*e.g.* root penetration, frost/thaw cycles; see Fig. 3.21C). This would possibly cause their hydraulic conductivity to be significantly higher than that

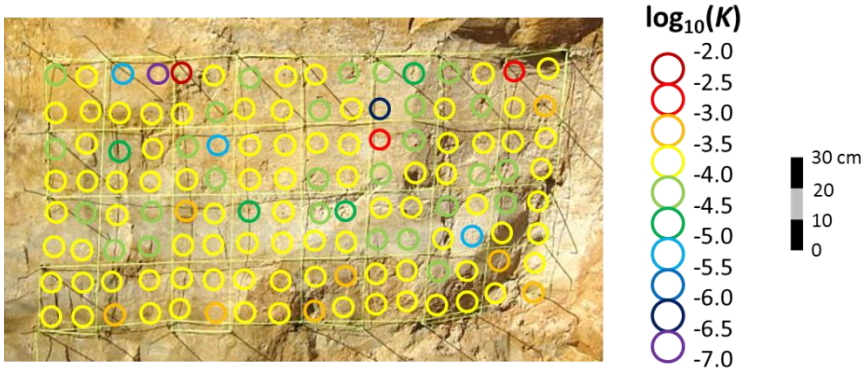

Figure 3.20. Stuctureless Kasterlee Sands outcrop.

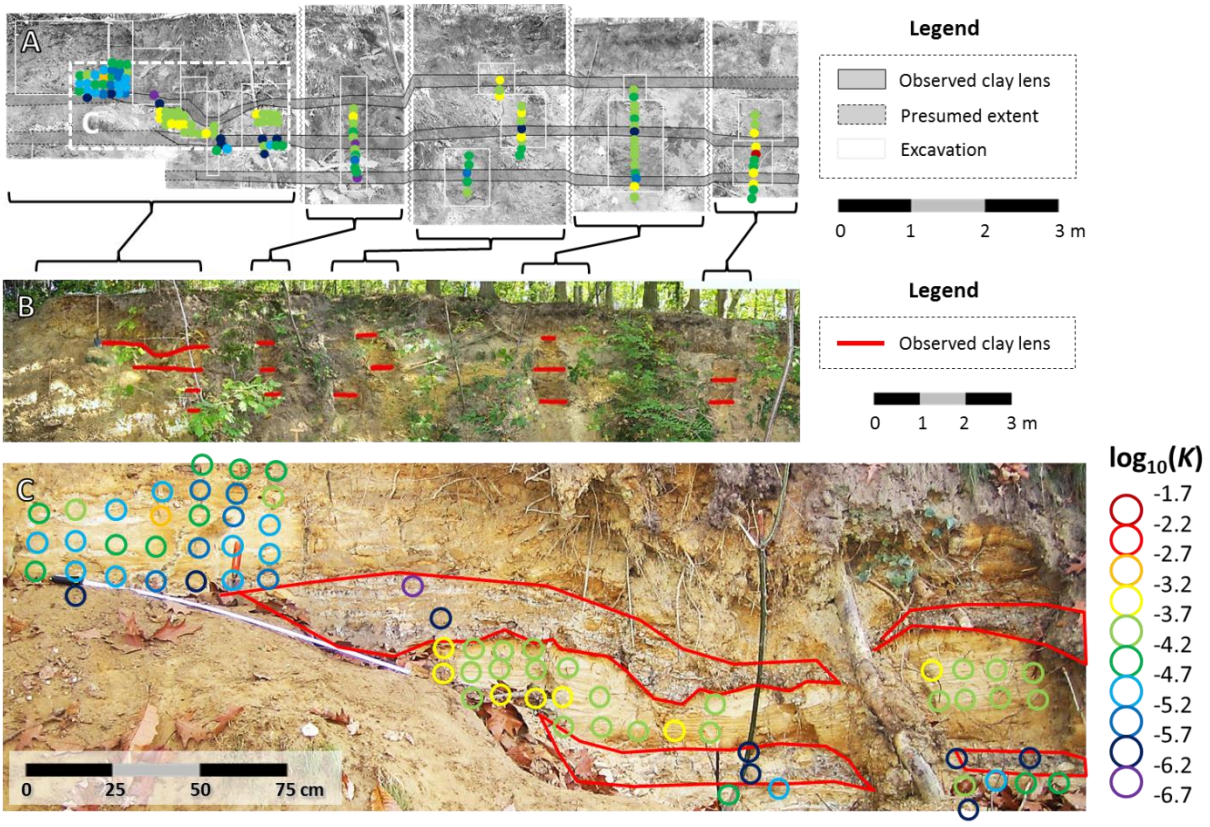

Figure 3.21. Kasterlee Clay outcrop at Heist-op-den-berg. A: Excavated profiles with clay lens connectivity. B: Overview of the entire outcrop with observed clay lenses. C: Detail of clay lenses with root penetration causing weathering of the clay.

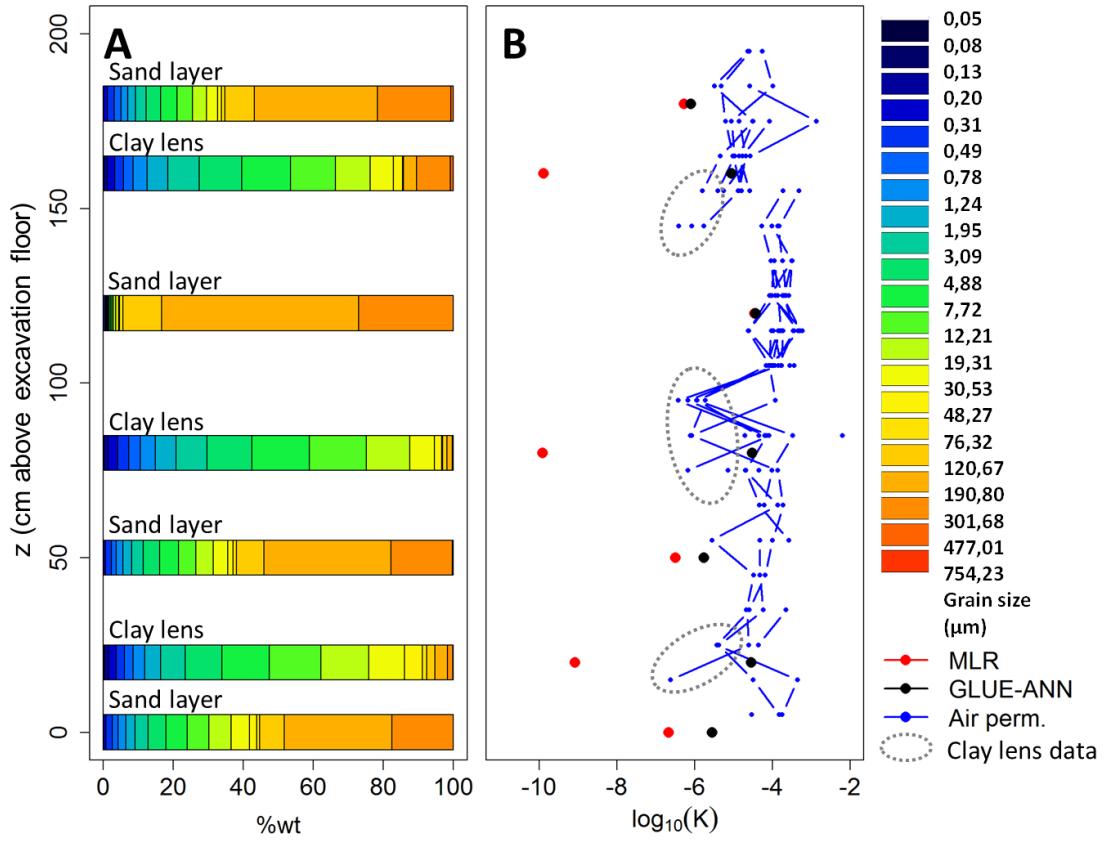

Figure 3.22. Grain size distributions (A) and hydraulic conductivities (B) derived from the air permeameter measurements (Air perm.), and predictions based on the multiple linear regression model (MLR) and GLUE-ANN ensemble of Rogiers *et al.* (2012; Chapter 6) for the Kasterlee Clay outcrop at Heist-op-den-berg. One grain size sample was taken within each lithological unit. The connected dots represent vertical transects.

of similar unweathered layers, and therefore the outcrop-scale hydraulic conductivity anisotropy would be smaller than that of unweathered layers such as those of the Kasterlee Clay aquitard.

In the left part of the outcrop, measurements were performed each 10 cm, while in the middle and right section, four vertical profiles were sampled. Measurements were performed on the clay lenses and the intercalated sand layers. The lowest *K* values were observed in the clay, ~ $10^{-6.6}$ m/s (Table 3.2), which is still relatively high given the high clay content (20%) and low sand content (< 15%; see Fig. 3.22). This can probably be explained by the fact that the clay is heavily weathered, also resulting in less variability and higher values for the clay than the bottom of the Poederlee Formation (outcrop g) and the transition to the underlying Kasterlee Sands (outcrop i; Fig. 3.5). Remarkably, the sand banks separated by these clay lenses showed clear differences in *K*. The upper one shows a clay content of 10% and silt content of 20% (Fig. 3.22), while *K* ranges from 10^{-4} to $10^{-5.5}$ m/s. The second sand bank has a sand content of more than 90%, and shows higher hydraulic conductivity values, between 10^{-3} and 10^{-4} m/s. The third sand bank is again similar to the first one, both in terms of hydraulic conductivity and grain size, as is the fourth one. The differences between these sand banks however seem to be less pronounced towards the right part of the outcrop, indicating that the horizontal correlation length within the sand might be at least several meters. The horizontal experimental variogram indeed shows a steady increase in variability with distance, but for the vertical one a pure nugget is present (Fig. 3.6).

To provide a full matrix of hydraulic conductivity values for the upscaling calculations for the Kasterlee Clay outcrop, a variogram model was fitted to the measurements within the sands only. Next, a geostatistical simulation was performed to fill in the sands,

while a constant value was used for the thick clay lenses. This is justified since the weathered state of the clay made it difficult to obtain reliable measurements. Therefore the lowest value was used. The heterogeneity within the sands determines the horizontal *K* value, and given the limited amount of data, this geostatistical approach has been applied. For more details, the reader is referred to Rogiers *et al.* (2013a; Chapter 2). The results of this simulation are shown in Fig. 3.4, and represent the entire outcrop section displayed in Fig. 3.21B. These are also used to derive outcrop-scale parameter values in section 3.4.

The grain size analyses for the clay lenses and the sand banks in between are shown in Fig. 3.22. The clay lenses have a clay content up to 20%, more than 70% of silt and only a few percent of sand. The sand banks have less silt (< 45%) and clay (< 10%); different sand banks however have a different particle size distribution. The difference between grain size and air permeameter based hydraulic conductivity values is again very clear. Also for the more sandy material, a consistent offset exists. The high fine silt contents could be the reason for this, and definitely causes the GLUE-ANN model to provide erroneous predictions.

3.3.3.2. The Diest Formation

The Diest Formation consists of grey green to brownish, mostly coarse, locally clayey, glauconiferous sand, often with sandstone layers (Laga *et al.*, 2001). Glauconite content can be as high as 50 wt% (Beerten *et al.*, 2010). The Diest Sands were deposited as sandbanks migrating over an erosive surface, and were influenced by coast-parallel tidal currents (Wouters & Vandenberghe, 1994). Paleontological evidence suggests that prolonged marginal marine conditions must have prevailed during the Late Miocene (Louwye & Laga, 2008).

Outcrop l: Diest Clayey Top at Heist-op-den-berg

The Heist-op-den-berg outcrop contains only the upper part of the Diest Formation (Fig. 3.23), referred to as Diest Clayey Top by Beerten *et al.* (2010). This is a relevant analogue layer as it was previously observed in nearly all aquifer boreholes used to develop the hydrogeological model for the Neogene aquifer (Beerten *et al.*, 2010). In all boreholes the Diest Clayey Top was overlain by the Kasterlee Clay. At its bottom, the outcrop consists of an alternation of sands with some darker-coloured clayey zones. At its top we observed an oxidized horizon containing concretions overlain by coarse sands which in turn were covered by more clayey material. The coarse sands are believed to be a weakly developed basal gravel corresponding to the base of the Kasterlee Formation (Fobe, 1995).

Air permeameter measurements were performed at 5 cm spacing because of the more visually distinguishable heterogeneity. The clayey zones, together with the coarse sand, are clearly the most important features determining hydraulic conductivity heterogeneity. Hydraulic conductivity for the clayey zones ranges between 10^{-4} and 10^{-6} m/s. For the sand in between the clayey zones, values are higher than $10^{-4.5}$ m/s, while the coarse sand shows values higher than $10^{-3.5}$ m/s (Fig. 3.4). The range is similar to that of the Kasterlee Clay outcrop, but the *K* values are higher by a factor five (Table 3.2). Spatial correlation is clearly present in the horizontal direction, but is less apparent in the vertical (Fig. 3.6, outcrop l). The entire measurement grid is used to calculate outcrop-scale parameter values in section 3.4.

Grain size analyses performed on the clayey zones and the typical sands (Fig. 3.24) illustrate a transition from a generally fine to slightly coarser material, with pronounced heterogeneity in the entire profile, as shown by the air permeability *K* estimates. The systematic offset between the grain size based predictions is again present in the entire outcrop.

Outcrop m: Diest Sands at Lummen

The outcrop of the Diest Sands in Lummen is situated on a remnant hill resisting erosion because of iron sandstones at the top of the sediments. The studied outcrop exhibits typical bioturbation structures and high glauconite content and shows several oxidized zones typical of the iron rich sands. A few thin clay lenses or clay drapes are present as well (Fig. 3.25).

The air permeameter measurements show a mean *K* value of $10^{-4.1}$ m/s (Table 3.2). The *K* range exceeds one order of magnitude, making it more heterogeneous than the Mol Sands, but less than the Poederlee and Kasterlee Sands (Fig. 3.5).

The spatial variability clearly increases with distance, reaching a higher sill level in the vertical (~0.2) than horizontal (~0.08) direction

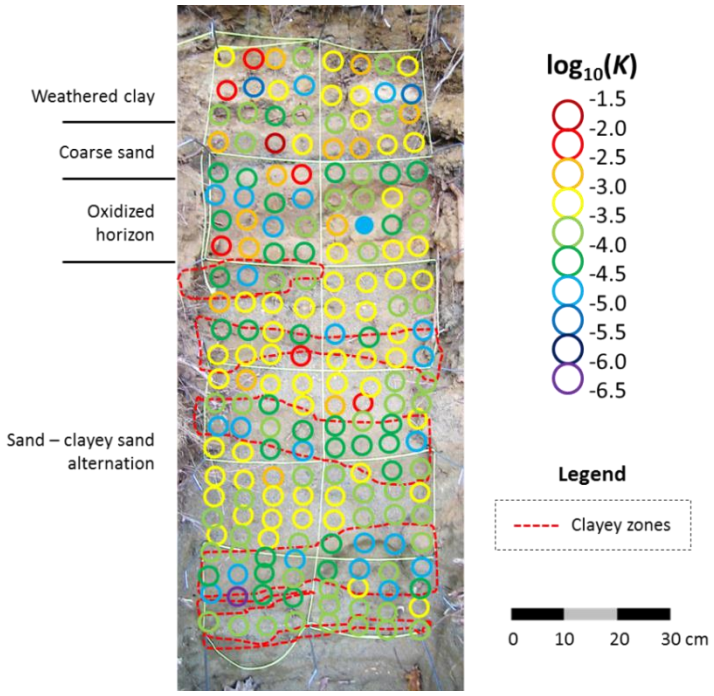

Figure 3.23. Diest Clayey Top outcrop.

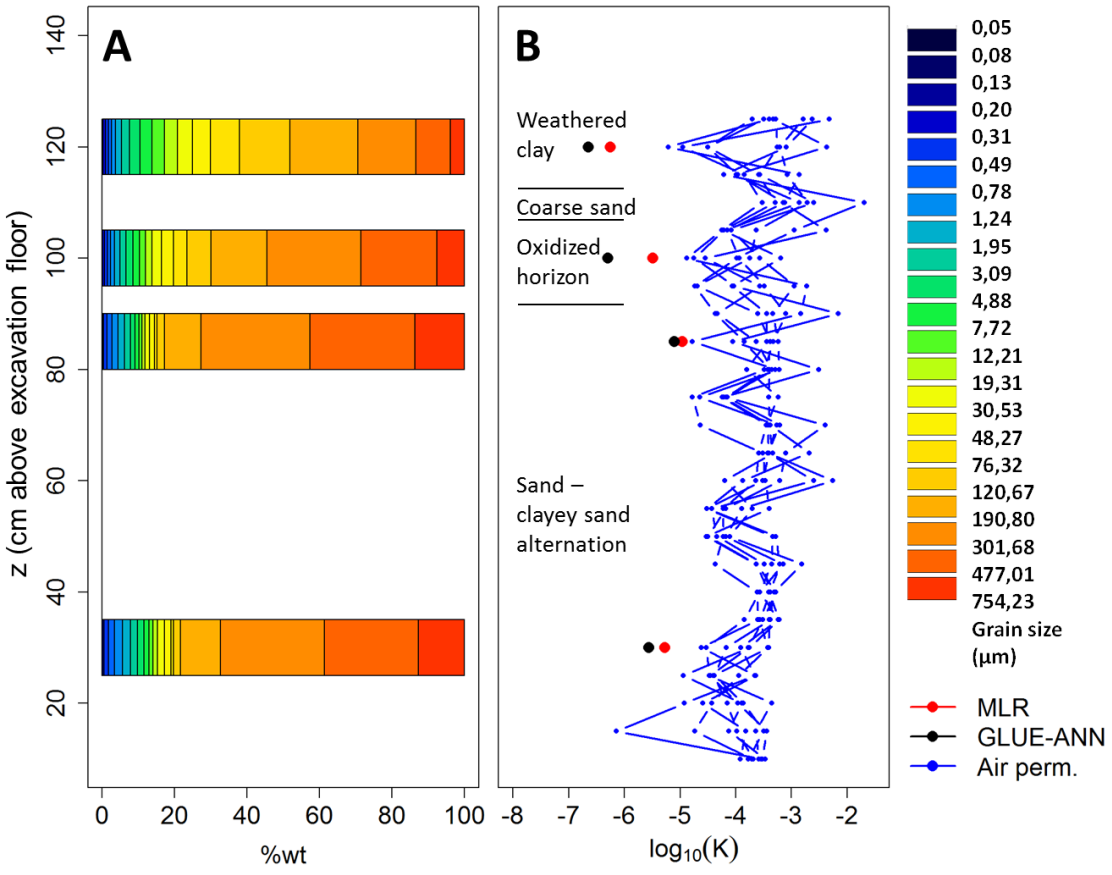

Figure 3.24. Grain size distributions (A) and hydraulic conductivities (B) derived from the air permeameter measurements (Air perm.), and predicted with the multiple linear regression model (MLR) and GLUE-ANN ensemble of Rogiers *et al.* (2012; Chapter 6) for Diest Clayey Top outcrop. The connected dots represent vertical transects.

(Fig. 3.6, outcrop m). The sampling density and the small measurement support did not allow including the very thin clay drapes in the analyses. Due to their limited horizontal continuity, however, they are probably not that important at larger scales. The contrast with the very heterogeneous Diest Clayey Top outcrop is not as expected, and it is doubtful that these outcrops are representative of the respective lithological units. The entire measurement grid is used to calculate outcrop-scale parameter values in section 3.4.

3.3.3.3. The Bolderberg Formation

The Bolderberg Formation is a lateral succession of marine to continental sandy deposits (Laga *et al.*, 2001). The white, fairly coarse sands with lignite layers and glassy quartzite banks, referred-to as Genk member, are studied here. Both outcrops (Figs 3.26 & 3.27) are located in the Sibelco quarry at Maasmechelen, where several meters of Meuse gravels overlie the Bolderberg Formation.

Outcrop n: Bolderberg Sands at Maasmechelen (upper part)

The upper outcrop (Fig. 3.26) is located only a few meters below the Meuse gravels, and shows a few layers of finer material, and some oxidized patches with higher iron content. Within the white sand, hardly any structures are distinguishable.

This Genk member of the Bolderberg Sands clearly is one of the most homogeneous sediments studied. Comparable to the Mol and Kleine Nete point bar sands, the K range is less than one order of magnitude (Fig. 3.5). The K values (mean value of $10^{-3.9}$ m/s) are also comparable with those from the Kleine Nete and Mol Sands (respectively outcrop a and e) (Table 3.2). The experimental variogram indicates the sampling grid was too small to capture the horizontal variability (no sill), while for the vertical direction the sill is reached after about 50 cm (Fig. 3.6, outcrop n). The entire measurement grid is used to calculate outcrop-scale parameter values in section 3.4.

Outcrop o: Bolderberg Sands at Maasmechelen (lower part)

The second Bolderberg Sands outcrop is even more homogeneous and shows no structures at all (Fig. 3.27). It is located only a few meters below the previous outcrop. Field observations indicated the sand at the top of the outcrop was finer than the sand below, explaining the trend in hydraulic conductivity.

The homogeneity is also clearly visible on the 10-cm spaced measurement grid (Fig. 3.4, outcrop o). As a result the $log_{10}(K)$ standard deviation is the overall smallest (0.15) across all outcrops (Table 3.2). The horizontal experimental variogram shows a pure nugget effect, but due to the trend in grain size the vertical variogram shows clear spatial correlation without a sill (Fig. 3.6, outcrop o). The entire measurement grid is used to calculate outcrop-scale parameter values in section 3.4.

3.4. Outcrop-scale equivalent K values and dispersivities

The equivalent K values obtained are listed in Table 2.3, both for the permeameter-type boundary conditions and the four alternative flow directions. The non-zero components of the K tensor are given together with the vertical anisotropy factor VANI.

The horizontal equivalent K values are all between $10^{-3.5}$ and $10^{-4.9}$ m/s. Hence, the variability of equivalent horizontal K within the studied outcrops of the Neogene aquifer remains limited. The vertical equivalent K values show more variations, ranging from $10^{-3.6}$ to $10^{-6.2}$ m/s. The relative differences between horizontal and vertical equivalent K, and between the different outcrops are consistent with our expectations based on K analysis on borehole cores (Beerten *et al.*, 2010), but the Diest Clayey Top exhibits higher conductivities relative to the other outcrops than expected. The four combined fixed head boundary conditions gave similar results to those from the permeameter-type setups. The off-diagonal components of the K tensor are all at least one order of magnitude smaller than the diagonal components, which justifies the use of diagonal-tensor models like *e.g.* MODFLOW (Harbaugh, 2005) in the large-scale simulations of the Neogene aquifer (see Gedeon *et al.*, 2011).

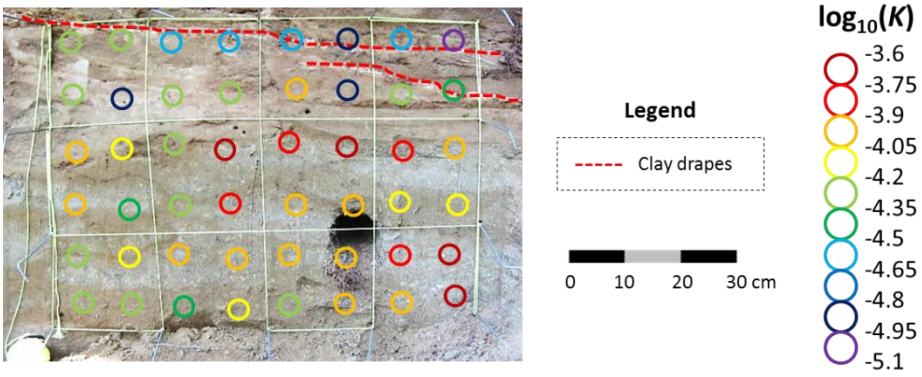

Figure 3.25. Diest Sands outcrop with thin clay drapes, bioturbations throughout the lower half of the outcrop, and oxidized patches.

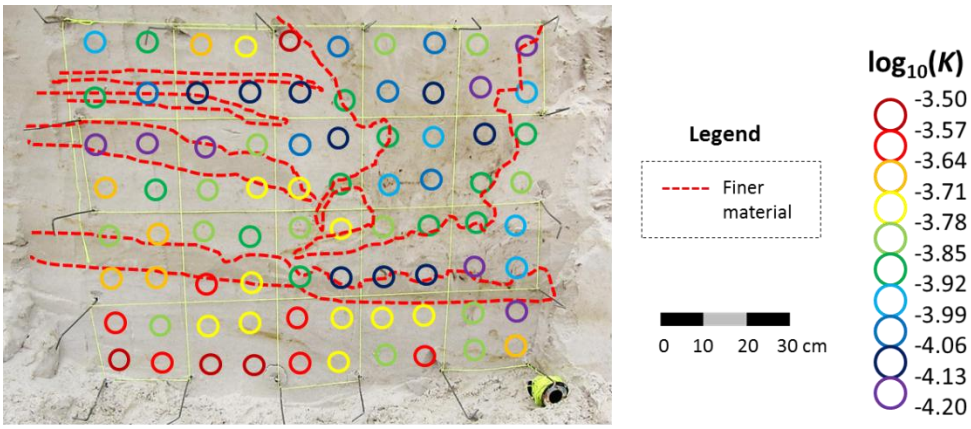

Figure 3.26. Upper Bolderberg Sands outcrop showing bands of finer material and oxidized patches.

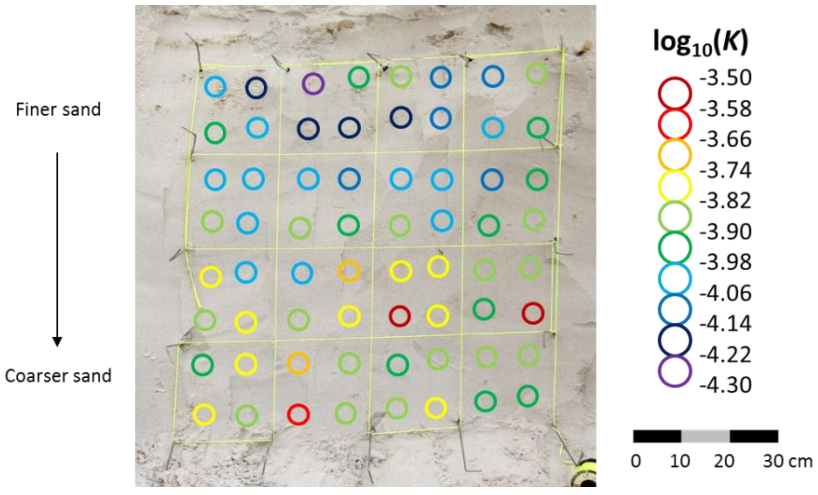

Figure 3.27. Very homogeneous structure-less Lower Bolderberg Sands outcrop characterized by a fining upwards trend.

Table 3.3. Overview of upscaled equivalent *K* values and dispersivities for the different grids presented in Figure 3.4.

	-Permeameter setup-			*-Combination of boundary conditions-*				*Horizontal flow*		*Vertical flow*	
ID	$log_{10}(K_{xx})$	$log_{10}(K_{zz})$	*VANI*	$log_{10}(K_{xx})$	$log_{10}(K_{zz})$	$log_{10}(K_{xz})$	*VANI*	$\alpha_L(m)$	$\alpha_{TV}(m)$	$\alpha_L(m)$	$\alpha_{TH}(m)$
b	-4.34**	-7.29**	890**	-	-	-	-	-	-	-	-
c	-4.15	-5.57	26.42	-4.16	-4.42	-9.66	1.82	*2.7E+02**	*4.3E-03**	*8.9E-01**	*1.3E-02**
d	-4.19	-4.19	1	-4.16	-4.24	-5.24	1.2	2.20E-01	3.50E-04	8.50E-02	3.00E-03
e	-3.64	-3.6	0.92	-3.68	-3.69	-5.37	1.03	3.50E-02	1.20E-04	1.70E-02	4.80E-04
f	-3.88	-5.31	26.71	-3.98	-4.79	-6.14	6.4	*7.1E-01**	*6.8E-04**	*1.8E-02**	*6.1E-05**
g	-4.8	-6.23	26.86	-4.86	-5.55	-12.59	4.97	*2.4E+00**	*1.9E-03**	*5.3E-02**	*2.8E-04**
h	-3.84	-4.02	1.51	-3.89	-3.97	-5.09	1.22	9.70E-02	8.00E-04	3.00E-02	2.50E-03
i	-	-6.05	-	-	-	-	-	-*	-*	-*	-*
j	-3.86	-3.85	0.98	-3.86	-3.87	-5.62	1.02	1.10E-01	1.00E-03	2.20E-02	2.10E-03
k	-4.13	-5.71	38.31	-4.05	-5.47	-13.08	26.46	*3.7E+09**	*2.1E-03**	*2.4E-02**	*1.8E-03**
l	-3.56	-3.88	2.08	-3.58	-3.77	-4.96	1.53	1.50E-01	1.40E-03	5.30E-02	3.40E-04
m	-4.06	-4.14	1.21	-4.09	-4.18	-12.23	1.22	7.90E-02	4.40E-04	1.20E-02	1.20E-03
n	-3.81	-3.84	1.07	-3.84	-3.86	-5.4	1.05	5.00E-02	3.50E-04	7.70E-03	2.00E-04
o	-3.86	-3.88	1.05	-3.9	-3.92	-8.61	1.04	2.60E-02	1.10E-04	2.80E-03	4.00E-04

* Presence of contrasting materials and structured heterogeneity, resulting in meaningless estimates of Fickian dispersivity.

** Calculated using arithmetic and harmonic means.

It should be noted here that the derived equivalent 2D *K* tensors are only representative for 3D *K*, in which the third dimension is constant, as outlined in Chapter 2. As this is not the case in reality, a discrepancy between the derived 2D and the true 3D values can be expected. For isotropic media, using the variance of $\log_{10}K$ to correct for the discrepancy between 2D and 3D using the results of Gutjahr *et al.* (1978) would lead to an increase in K by at most 22%, while in most cases the increase would be below 10% for the outcrops discussed here. Given the lack of clear correlation lengths in all different cases, we can only make a rough estimation for anisotropic media using the generalization of Matheron's conjecture (Matheron 1967) by Ababou (1996). With a horizontal correlation length of 2 m and a vertical one of 0.5 m, which seems reasonable at the outcrop scale given the empirical semivariograms in Fig. 3.6, the max discrepancy between 2D and 3D would amount 28% for K_v. For larger anisotropies, the differences only decrease.

Concerning the calculated vertical anisotropy parameter VANI, we conclude that the Mol, Kasterlee and Bolderberg Sands remain isotropic at the scale of the numerical grid, while the Lommel Sands show a slight anisotropy with a maximum VANI of 1.20. For the Poederlee Sands (at Lichtaart), Diest Sands, and Diest Clayey Top outcrops, the VANI parameter is between 1.2 and 2. Higher values are obtained for the Kasterlee Clay, Poederlee Sands (at Kasterlee), and Campine Complex outcrop with frost structures, ranging from 26 to 40. Finally, the profile through the Campine Complex produces the overall highest value for VANI equal to 890.

PMPATH (included in Processing Modflow; Chiang and Kinzelback, 2001) was used to perform the particle tracking calculations. It uses a semi-analytical scheme (Pollock, 1988) and as input the simulation results from the flow model to calculate the groundwater paths and travel times. At the inflow boundary, 10 particles were inserted per grid cell to get a representative distribution.

The breakthrough curves of many of these particle tracking simulations show heavy tailing and multiple peaks (not shown). This indicates that a Fickian dispersion model (Bear, 1979) would not be appropriate for these geologic media, as this model has several shortcomings (Konikow, 2011), or that the representative elementary volumes of the studied sediments exceed the outcrop sizes. Nevertheless, dispersivities were estimated and the results presented in Table 2.3. The outcrops with distinctly different and contrasting materials with structured heterogeneity are marked with an asterisk, as their calculated dispersivities are very large and hence not meaningful. All other longitudinal dispersivities are between 10^{-3} and 10^{0} m, and their transversal equivalents are one to three orders of magnitude smaller.

Fig. 3.28 shows the longitudinal dispersivity as a function of measurement scale. Only the meaningful values (Table 3.3) are included in the graph. Gelhar *et al.* (1985, 1992) and Schulze-Makuch (2005) collected published dispersivity data to estimate the scaling relationship. The work by Baumann *et al.* (2002, 2005, 2010) is focussed on pore-scale dispersion, and provides results that are in line with the macro-dispersion parameters obtained in the field. Kim *et al.* (2002) used TDR measurements to determine at the centimetre scale the longitudinal dispersivity in an unconfined sandy aquifer, which complements the previous two types of data sets. Finally, the natural gradient tracer study by Mallants *et al.* (2000) provided several dispersivity values for the Brasschaat and Mol Sands of the Neogene aquifer; such field-based values are useful in a comparison with the numerically calculated values from the present study.

Based on Fig. 3.29 all available data display a high degree of consistency with the observed scatter being representative of different degrees of sediment heterogeneity. The values we obtained with this fairly simple dispersivity calculation approach are consistent with the literature data. The four outliers in Table 2.3 are the Kasterlee Clay, the Campine Complex frost structures outcrop, and both Poederlee outcrops at Kasterlee. Because of the highly contrasting lithologies in these outcrops, a dual domain approach (*e.g.* Flach *et al.*, 2004) might be more appropriate. The Mol and Bolderberg Sand outcrops show the lowest dispersivity values, although a difference of about one order of magnitude exists between the values obtained from considering horizontal and vertical flow directions; this indicates that at least an anisotropic dispersion model should be used. These low dispersivity values also accord most with the data obtained from similar sediments by Mallants *et al.* (2000).

Vertical and horizontal transverse dispersivities are shown in Fig. 3.29, again without the outliers from Table 2.3. The data from Gelhar *et al.* (1992) shows α_L/α_T ratios between 1 and 100 for the horizontal transverse dispersivity (α_{TH}), and up to 1000 for the vertical one (α_{TV}). This is consistent with our data, although the effect of flow direction (*i.e.* horizontal versus vertical) on the α_L/α_T ratios is clearly visible.

These analyses of dispersivities demonstrate that the use of isotropic dispersion coefficients in such highly stratified sediments may not yield correct results; approaches that account for effects of layering on dispersion are preferred. For the few sediments with highly contrasting materials and structured heterogeneity more complex approaches are recommended (such as dual domain approaches or more explicit representation of the small-scale heterogeneity instead of using a macro-description).

3.5. Conclusions

Air permeameter measurements are an efficient and flexible way to characterize hydraulic properties of outcropping sediments. Investigations on different lithologies, vertical profiles, or even two-dimensional regular grids are generally straightforward. Besides effectively characterizing hydraulic conductivity and its small-scale spatial variability, numerical methods can be used in conjunction to derive an equivalent *K* tensor at the scale of the entire outcrop, as well as the dispersivity transport parameter.

The geometric mean air permeameter based *K* values for the outcrops discussed in this paper range from $10^{-5.94}$ m/s to $10^{-3.68}$ m/s, whereas the minimum and maximum of the individual measurements amounted respectively to $10^{-9.17}$ and $10^{-1.56}$ m/s. The maximum difference between values obtained for one outcrop ranged from less than one order of magnitude to more than four orders of magnitude, indicating the presence of important small-scale heterogeneities as well as several very homogeneous outcrops.

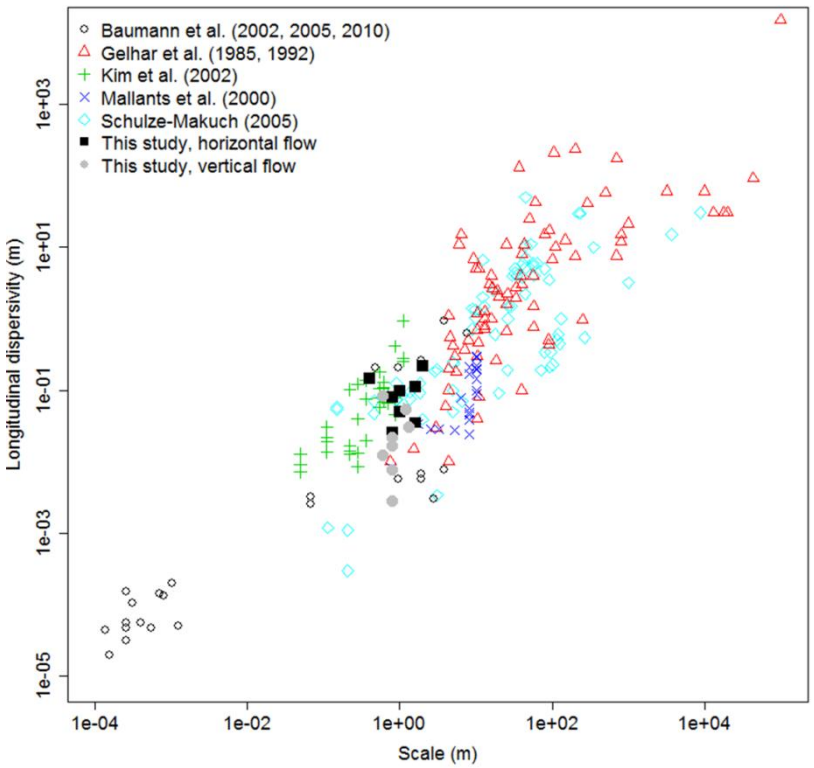

Figure 3.28. Longitudinal dispersivity as a function of measurement scale. Literature values are combined with numerically calculated values for outcrops. Data from Mallants *et al.* (2000) was also obtained from Neogene aquifer sediments.

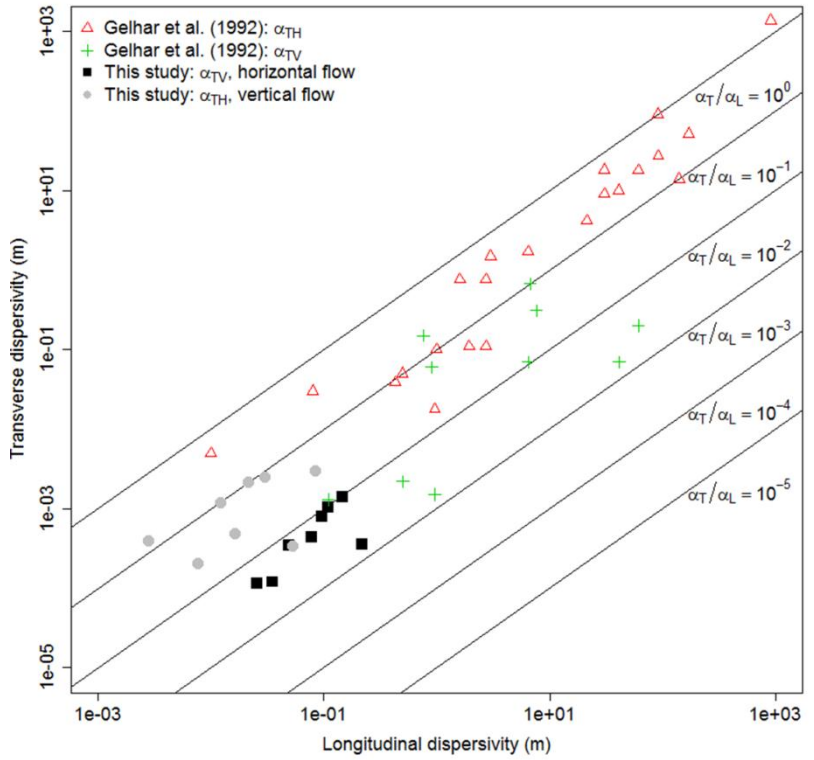

Figure 3.29. Numerically calculated longitudinal versus transverse dispersivity (both horizontal and vertical) for different flow directions combined with literature data from Gelhar *et al.* (1992).

Investigating spatial correlation for the different measurement grids showed that, at the considered outcrop scales, hardly any correlation exists for several formations (*e.g.* e, Mol Sands; g, bottom of Poederlee Fm.; j, Kasterlee Sands), while others display a significant increase of the variability (*i.e.* variance) with distance (*e.g.* i, Poederlee-Kasterlee transition zone; k, Kasterlee Clay; l, Diest Clayey Top; m, Diest Sands; n, Bolderberg Sands). Most horizontal experimental variograms do not reach the semivariance values of the corresponding vertical ones, which may indicate that the horizontal outcrop extents are only a fraction of the true horizontal correlation lengths, or, for a layered outcrop, that the variability within a layer is lower than variability across layers. Moreover, the contrasts in heterogeneity between the different formations are very well illustrated by means of the experimental variograms.

The grain size based *K* predictions confirmed the systematic offset between similar outcropping and subsurface sediments as discussed by Rogiers *et al.* (2013a; Chapter 2). The offset seems to be larger for the more clayey sediments that are also more affected by weathering. For recently excavated outcrops, with still several meters of overburden, the bias between grain size and air permeameter based predictions seemed to be negligible (Figure 3.16). Moreover, the discrepancies between *K* measurements and predictions based on the non-linear site-specific GLUE-ANN model (Rogiers *et al.*, 2012; Chapter 6) revealed that the discrepancies are highest where fine silt fractions exist in outcrop lithologies, a result of such textural fractions not occurring in the borehole data on which the model was trained.

The numerically upscaled equivalent *K* values obtained for the different outcrops range from $10^{-4.9}$ m/s to $10^{-3.6}$ m/s in the horizontal direction and from $10^{-6.2}$ m/s to $10^{-3.6}$ m/s in the vertical direction. A slight bias might however exist between these 2D-derived values and 3D values. The corresponding vertical anisotropy values range between 1 and 40. Most of the calculated longitudinal dispersivities fit well within literature data, given the scale of the outcrops. A few outliers were obtained for outcrops with contrasting lithologies, where a dual domain approach might be more appropriate. The ratio between longitudinal and transverse dispersivities seems to be rather high, but falls in the range reported in literature. Moreover, the dispersivity values obtained by Mallants *et al.* (2000) at a slightly larger scale are very similar. This builds confidence for the use of numerically calculated dispersivity values from outcrop sediments within larger-scale contaminant transport models.

Increasing size and number of investigated outcrops of the same formations would provide more information on the larger-scale variability. A systematic study addressing *K*-dependency on sediment porosity, density, compaction and weathering processes might provide adequate tools to convert the outcrop *K*-data into effective values for subsurface groundwater flow and contaminant transport simulation.

References

Ababou, R., 1996. Random porous media flow on large 3D grids: numerics, performance, and application to homogenization, Chap 1, vol 79. In: Wheeler MF (ed) Environmental studies, mathematical, computational & statistical analysis. Springer, New York, pp 1–25

AMINAL, 2003. De Kleine Nete en de Aa - Naar een ecologisch herstel van waterloop en vallei (p. 26). Afdeling Water, Brussel.

Baumann, T., Müller, S., Niessner, R., 2002. Migration of dissolved heavy metal compounds and PCP in the presence of colloids through a heterogeneous calcareous gravel and a homogeneous quartz sand – pilot scale experiments. *Water Research* **36**: 1213–1223.

Baumann, T., Werth, C.J., 2005. Visualization of colloid transport through heterogeneous porous media using magnetic resonance imaging. *Colloids and Surfaces A: Physicochemical and Engineering Aspects* **265**: 2–10.

Baumann, T., Toops, L., Niessner, R., 2010. Colloid dispersion on the pore scale. *Water Research* **44**(4): 1246-54.

Bear, J., 1979. Hydraulics of Groundwater, 569. New York: McGraw-Hill.

Beaudoin, A., de Dreuzy, J.-R., 2013. Numerical assessment of 3-D macrodispersion in heterogeneous porous media. *Water Resources Research* **49**(5): 2489-2496.

Beerten, K., 2006. Toelichting bij de Quartairgeologische kaart, kaartblad 17 Mol. Vlaamse Overheid, Dienst Natuurlijke Rijkdommen.

Beerten, K., 2010. Geomorphological evolution of the Nete basin: identification of past events to assess the future evolution. SCK•CEN external report (p. 41). Mol.

Beerten, K., Wemaere, I., Gedeon, M., Labat, S., Rogiers, B., Mallants, D., Salah, S., Leterme, B.,

2010. Geological, hydrogeological and hydrological data for the Dessel disposal site. Project near surface disposal of category A waste at Dessel – Version 1.- Brussels, Belgium: NIRAS/ONDRAF, 2010.- 273 p.- NIROND–TR 2009–05 E V1.

Chiang, W., Kinzelbach, W., 2001. 3D-groundwater modeling with PMWIN. Springer, Berlin. ISBN 3-540- 67744-5

Coetsiers, M., Walraevens, K., 2006. Chemical characterization of the Neogene Aquifer, Belgium. *Hydrogeology Journal* **14**: 1556-1568.

Coetsiers, M., Blaser, P., Martens, K., Walraevens, K., 2008. Natural background levels and threshold values for groundwater in fluvial Pleistocene and Tertiary marine aquifers in Flanders, Belgium. *Environmental Geology* **57**(5): 1155-1168.

De Bie, E., Martens, K., Haesevoets, A., Florus, M., 2007. Rivierherstel van de Kleine Nete tussen Herentals en Kasterlee. Water, Congres Watersysteemkennis 2006-2007, Bodem, Grondwater en Ecosysteem (pp. 1-5).

De Marsily, G., Delay, F., Goncalves, J., Renard, P., Teles, V., Violette, S., 2005. Dealing with spatial heterogeneity. *Hydrogeology Journal* **13**(1): 161–183.

Flach, G. P., Crisman, S.A., Molz, F. J., 2004. Comparison of single-domain and dual-domain subsurface transport models. *Ground Water* **42**(6-7): 815–28.

Fobe, B., 1995. Litologie en litostratigrafie van de Formatie van Kasterlee (Plioceen van de Kempen). Natuurwetenschappelijk Tijdschrift, 75(2): 35-45.

Gedeon, M., Wemaere, I., Marivoet, J., 2007. Regional groundwater model of north-east Belgium. *Journal of Hydrology* **335**(1-2): 133-139.

Gedeon, M., 2008. Neogene Aquifer Model.- Mol, Belgium: SCK•CEN, 2008.- 110 p.- (External Report of the Belgian Nuclear Research Centre; ER-48; CCHO-2004-2470/00/00, DS 251-A51).- ISSN 1782-2335

Gedeon, M., Mallants, D., Vandersteen, K., Rogiers, B., 2011. Hydrogeological modelling of the Dessel site - Overview report. Project near surface disposal of category A waste at Dessel. NIROND-TR report 2008–15 E V2

Gedeon, M., Mallants, D., 2012. Sensitivity Analysis of a Combined Groundwater Flow and Solute Transport Model Using Local-Grid Refinement: A Case Study. *Mathematical Geosciences* **44**(7): 881-899.

Gelhar, L.W., Mantoglou, A., Welty, C., Rehfeldt, K.R., 1985. A review of field-scale physical solute transport processes in saturated and unsaturated porous media. Palo Alto, California: Electric Power Research Institute EPRI EA-4190 Project 2485–5.

Gelhar, L.W., Welty, C., Rehfeldt, K., 1992. A Critical Review of Data on Field-Scale Dispersion in Aquifers. *Water Resources Research* **28**(7): 1955-1974.

Gelhar, Lynn W., Stochastic Subsurface Hydrology, Prentice Hall, 390 pp., 1993.

Gullentops, F., Wouters, L. (Eds.), 1996. Delfstoffen in Vlaanderen. Ministerie Vlaamse Gemeenschap, Brussel, 198pp.

Gullentops, F., Bogemans, F., De Moor, G., Paulissen, E., Pissart, A., 2001. Quaternary lithostratigraphic units (Belgium). *Geologica Belgica* **4**(1-2): 153–164.

Gutjahr, A.L., Gelhar, L.W., Bakr, A.A., McMillan, J.R., 1978. Stochastic analysis of spatial variability in subsurface flow 2: Evaluation and application. *Water Resour. Res.* **14**: 953–959

Harbaugh, A.W., 2005. MODFLOW-2005, The U.S. Geological Survey modular ground-water model—the Ground-Water Flow Process: U.S. Geological Survey Techniques and Methods 6-A16, USGS, 2005

Huysmans, M., Peeters, L., Moermans, G., Dassargues, A., 2008. Relating small-scale sedimentary structures and permeability in a cross-bedded aquifer. *Journal of Hydrology* **361**(1-2): 41-51.

Huysmans, M., Dassargues, A., 2009. Application of multiple-point geostatistics on modelling groundwater flow and transport in a cross-bedded aquifer (Belgium). *Hydrogeology Journal* **17**(8): 1901–1911.

Iversen, B.V., Moldrup, P., Schjonning, P. & Jacobsen, O.H., 2003. Field Application of a Portable Air Permeameter to Characterize Spatial Variability in Air and Water Permeability. *Vadose Zone Journal* **2**(4): 618-626.

Jacques, D., Leterme, B., Beerten, K., Schneider, S., Finke, P., Mallants, D., 2010. Long-term evolution of the multi-layer cover. Project near surface disposal of category A waste at Dessel. NIROND-TR 2010-03 E, September 2010.

Kim, D.-J., Kim, J.-S., Yun, S.-T., Lee, S.-H., 2002. Determination of longitudinal dispersivity in an unconfined sandy aquifer. *Hydrological Processes* **16**(10): 1955–1964.

Koltermann, C., Gorelick, S., 1996. Heterogeneity in Sedimentary Deposits: A Review of Structure Imitating, Process Imitating, and Descriptive Approaches. *Water Resources Research* **32**(9): 2617-2658.

Konikow, L.F., 2011. The secret to successful solute-transport modeling. *Ground Water* **49**(2): 144-59.

Laga, P., Louwye, S., Geets, S., 2001. Paleogene and Neogene lithostratigraphic units (Belgium). *Geologica Belgica* **4**(1-2): 135–152.

Li, L., Zhou, H., Jaime Gómez-Hernández, J., 2011. A Comparative Study of Three-Dimensional Hydraulic Conductivity Upscaling at the MAcro-Dispersion Experiment (MADE) site, Columbus

Air Force Base, Mississippi (USA). *Journal of Hydrology* **404**(3-4): 278-293.

Louwye, S., De Schepper, S., Laga, P., Vandenberghe, N., 2007. The Upper Miocene of the southern North Sea Basin (northern Belgium): a palaeoenvironmental and stratigraphical reconstruction using dinoflagellate cysts. *Geological Magazine* **144**(1): 33.

Louwye, S., Laga, P., 2008. Dinoflagellate cyst stratigraphy and palaeoenvironment of the marginal marine Middle and Upper Miocene of the eastern Campine area, northern Belgium (southern North Sea Basin). *Geological Journal* **43**(1): 75–94.

Louwye, S., De Schepper, S., 2010. The Miocene–Pliocene hiatus in the southern North Sea Basin (northern Belgium) revealed by dinoflagellate cysts. *Geological Magazine* **147**(05): 760-776.

Mallants, D., Espino, A., Hoorick, M., Feyen, J., Vandenberghe, N. & Loy, W., 2000. Dispersivity estimates from a tracer experiment in a sandy aquifer. *Ground Water* **38**(2) : 304–310.

Matheron G, 1967. Éléments pour une théorie des milieux poreux. Masson and Cie, Paris

Meyus, Y., Batelaan, O., De Smedt, F., 2000. Concept Vlaams Grondwater Model (VGM). Deelrapport: Technisch concept van het VGM: Hydrogeologische Codering van de Ondergrond van Vlaanderen (HCOV). (p. 58). Brussel.

New England Research, Vindum Engineering, 2011. TinyPerm II Portable Air Permeameter, User's Manual. Retrieved from http://www.vindum.com/TinyPermManual.pdf on 14-06-2011.

ONDRAF/NIRAS, 2010. Het cAt-project in Dessel. Een langetermijnoplossing voor het Belgische categorie A-afval. Retrieved from http://www.niras-cat.be/downloads/cAt_masterplan_NL_LOW.pdf on 07-12-2011.

Patyn, J., Ledoux, E., Bonne, A., 1989. Geohydrological research in relation to radioactive waste disposal in an argillaceous formation. *Journal of Hydrology* **109**: 267–285.

Pollock, D.W., 1988. Semianalytical Computation of Path Lines for Finite-Difference Models. *Ground Water* **26**(6): 743-750.

Possemiers, M., Huysmans, M., Peeters, L., Batelaan, O., Dassargues, A., 2012. Relationship between sedimentary features and permeability at different scales in the Brussels Sands. *Geologica Belgica* **15**(3): 156-164.

Rogiers, B., Schiltz, M., Beerten, K., Gedeon, M., Mallants, D., Batelaan, O., Dassargues, A., Huysmans, M., 2010a. Groundwater model parameter identification using a combination of cone-penetration tests and borehole data. IAHR international groundwater symposium (p. 19). Valencia.

Rogiers, B., Mallants, D., Batelaan, O., Gedeon, M., Huysmans, M., Dassargues, A., 2010b. Caractérisation de l'hétérogénéité de la conductivité hydraulique à saturation au moyen d'essais de pénétration au cône. 35 èmes journées scientifiques du Groupe Francophone d'Humidimétrie et des TraNsferts en Milieux Poreux: Transferts en milieux poreux : Hétérogénéité des processus et des propriétés, GFHN 2010, Bulletin du G.F.H.N. N°56: 149-150. Louvain-la-Neuve, 23-25 November 2010. - ISSN: 0997 - 1076

Rogiers, B., Mallants, D., Batelaan, O., Gedeon, M., Huysmans, M., Dassargues, A., 2010c. Geostatistical analysis of primary and secondary data in a sandy aquifer at Mol/Dessel, Belgium. 8th International conference on Geostatistics for Environmental Applications, geoENV 2010, Gent, 13-15 September 2010.

Rogiers, B., Mallants, D., Batelaan, O., Gedeon, M., Huysmans, M., Dassargues, A., 2010d. Exploratory and structural data analysis of a sandy aquifer at Mol/Dessel, Belgium. *Geophysical Research Abstracts* **12**, EGU General Assembly 2010, Vienna, 3-7 May 2010.

Rogiers, B., Beerten, K., Smekens, T., Mallants, D., 2011. Air permeability measurements on Neogene and Quaternary sediments from the Campine area: using outcrop analogues for determining hydrodynamic aquifer properties. SCK•CEN ER-177 (p. 23). Mol.

Rogiers, B., Mallants, D., Batelaan, O., Gedeon, M., Huysmans, M., Dassargues, A., 2012. Estimation of hydraulic conductivity and its uncertainty from grain size data using GLUE and artificial neural networks. *Mathematical Geosciences* **44**: 739–763.

Rogiers, B., Beerten, K., Smeekens, T., Mallants, D., Gedeon, M., Huysmans, M., Batelaan, O., Dassargues, A., 2013a. The usefulness of outcrop analogue air permeameter measurements for analysing aquifer heterogeneity: Quantifying outcrop hydraulic conductivity and its spatial variability. *Hydrological Processes*. DOI: 10.1002/hyp.10007

Rogiers, B., Beerten, K., Smeekens, T., Mallants, D., Gedeon, M., Huysmans, M., Batelaan, O., Dassargues, A., 2013b. The usefulness of outcrop analogue air permeameter measurements for analysing aquifer heterogeneity: Testing outcrop hydrogeological parameters with independent borehole data. *Hydrol. Earth Syst. Sci. Discuss.* **10**: 9689-9720.

Ronayne, M.J., Gorelick, S.M., Zheng, C., 2010. Geological modeling of submeter scale heterogeneity and its influence on tracer transport in a fluvial aquifer. *Water Resources Research* **46**(10): 1-9.

Salamon, P., Fernàndez-Garcia, D., Gómez-Hernández, J.J., 2006. A review and numerical

assessment of the random walk particle tracking method. *Journal of Contaminant Hydrology* **87**(3-4): 277-305.

Schaap, M., Leij, F., Van Genuchten, M.Th., 2001. Rosetta: A computer program for estimating soil hydraulic parameters with hierarchical pedotransfer functions. *Journal of Hydrology* **251**(3–4): 163–176.

Schulze-Makuch, D., 2005. Longitudinal dispersivity data and implications for scaling behavior. *Ground Water* **43**(3): 443–456.

Sibelco, 2010. Silica sand of Mol. Technical Datasheet TDS.03.05.10 2010-12-09

Seuntjens, P., Mallants, D., Simunek, J., Patyn, J., Jacques, D., 2002. Sensitivity analysis of physical and chemical properties affecting field-scale cadmium transport in a heterogeneous soil profile. *Journal of Hydrology* **264**(1-4): 185-202.

Van Camp, M., Coetsiers, M., Martens, K., Walraevens, K., 2010. Effects of multi-annual climate variability on the hydrodynamic evolution (1833-present) in a shallow aquifer system in northern Belgium. *Hydrological Sciences Journal* **55**(5): 763-779.

Van Camp, M., Martens, K., Walraevens, K., 2012. Impact of recent climate variability on an aquifer system in north Belgium The Netherlands Campine Cuesta Nete Basin Campine Plateau. *Geologica Belgica* **15**(1-2): 73–80.

Vandenberghe, N., Van Simaeys, S., Steurbaut, E., Jagt, J.W.M. & Felder, P.J., 2004. Stratigraphic architecture of the Upper Cretaceous and Cenozoic along the southern border of the North Sea Basin in Belgium. *Netherlands Journal of Geosciences* **83**: 155-171.

Vienken, T., Dietrich, P., 2011. Field evaluation of methods for determining hydraulic conductivity from grain size data. *Journal of Hydrology* **400**(1-2): 58-71.

VMM, 2005. Grondwaterbeheer in Vlaanderen: het onzichtbare water doorgrond. Bijlage 4: Vergunde debieten per grondwaterlichaam. Retrieved from http://www.vmm.be/water/toestand-watersystemen/watersysteemkennis/watersysteemkennis-grondwater/rapporten-grondwaterbeheer-in-vlaanderen-het-onzichtbare-water-doorgrond on 21-02-2012.

VMM, 2008. Het Centraal Kempisch Systeem (p. 108).

Wouters, L., Vandenberghe, N., 1994. Geologie van de Kempen. Niras: Brussel.

Yu, L., Rogiers, B., Gedeon, M., Wemaere, I., Marivoet, J., De Craen, M., Mallants, D., 2012. A critical review of laboratory and in-situ hydraulic conductivity measurements for the Boom Clay in Belgium. *Applied Clay Science* **75-76**: 1-12.

Chapter 4

Testing outcrop hydrogeological parameters with independent borehole data

Based on Rogiers B, Beerten K, Smeekens T, Mallants D, Gedeon M, Huysmans M, Batelaan O, Dassargues A. 2013. The usefulness of outcrop analogue air permeameter measurements for analysing aquifer heterogeneity: testing outcrop hydrogeological parameters with independent borehole data. Hydrol. Earth Syst. Sci., accepted. (Hydrol. Earth Syst. Sci. Discuss. 10: 9689-9720)

Abstract

Outcropping sediments can be used as easily accessible analogues for studying subsurface sediments, especially to determine the small-scale spatial variability of hydrogeological parameters. The use of cost-effective *in situ* measurement techniques potentially makes the study of outcrop sediments even more attractive. We investigate to what degree air permeameter measurements on outcrops of unconsolidated sediments can be a proxy for aquifer saturated hydraulic conductivity (*K*) heterogeneity. The Neogene aquifer in northern Belgium, known as a major groundwater resource, is used as case study. *K* and grain size data obtained from different outcropping sediments are compared with *K* and grain size data from aquifer sediments obtained either via laboratory analyses on undisturbed borehole cores (*K* and grain size) or via large-scale pumping tests (*K* only). This comparison shows a pronounced and systematic difference between outcrop and aquifer sediments. Part of this difference is attributed to grain size variations and earth surface processes specific to outcrop environments, including root growth, bioturbation, and weathering. Moreover, palaeoenvironmental conditions such as freezing-drying cycles and differential compaction histories will further alter the initial hydrogeological properties of the outcrop sediments. A linear correction is developed for rescaling the outcrop data to the subsurface data. The spatial structure pertaining to outcrops complements that obtained from the borehole cores in several cases. The higher spatial resolution of the outcrop measurements identifies small-scale spatial structures that remain undetected in the lower resolution borehole data. Insights in stratigraphic and *K* heterogeneity obtained from outcrop sediments improve developing conceptual models of groundwater flow and transport.

4.1. Introduction

Compared to core drilling for sample collection and analysis, outcropping sediments are easily accessible analogues for studying subsurface sediments. This outcrop analogue concept has been extensively applied in the oil industry for the analysis and modelling of reservoirs (*e.g.* Flint and Bryant, 1993; McKinley *et al.*, 2004) resulting in various tools to characterize geological facies geometries, their connectivity and continuity (Pringle *et al.*, 2004), and to create 3D virtual outcrop models (Pringle *et al.*, 2006). The concept has also been used with small-scale outcrops in unconsolidated material (*e.g.* Teutsch *et al.*, 1998, Bayer *et al.*, 2011), collecting both hydraulic and geophysical data. Most of these studies are more concerned with defining the geological facies geometry rather than determining the corresponding hydrogeological parameters and hence direct quantification of these parameters and certainly a comparison with the corresponding subsurface parameters is often lacking.

In slightly dipping unconsolidated stratigraphic settings, a very limited number of facies are generally encountered in a single outcrop. The information contained within such lithofacies type potentially represents key stratigraphic features and hydrogeological parameters for building conceptual groundwater flow models. Furthermore, different outcrops may represent different parts of a stratigraphic or landscape

succession series (Beerten *et al.*, 2012). The combination of several outcrops can then be used to obtain a composite picture of an aquifer system containing the same or at least similar sediments. As demonstrated by Rogiers *et al.* (2013a; Chapter 2), the use of a hand-held air permeameter is a very accurate and cost-effective approach for quantifying hydraulic conductivity (*K*) and its spatial variability *in situ* on outcropping sediments. The question that remains however is how representative the obtained outcrop parameters are for the actual subsurface sediments.

In first instance, the outcrop sediments may differ in some aspects from their subsurface equivalents as a result of slightly differing depositional contexts, *e.g.* with respect to the position in the basin (palaeogeographical conditions).

Secondly, the outcropping sediments could also be influenced by post-depositional processes such as glauconite disintegration or oxidation, surficial weathering and compaction due to slightly different overburden sedimentation and erosion histories. During the initial loading of sands, a rapid increase of packing density and soil strength is expected due to grain reorganization (Pettersen, 2007). As packing becomes tighter, further packing will be increasingly more difficult to achieve, each packing level is more stable than previous levels and deformation is permanent. This process should be visible in the porosity, bulk density and eventually *K* data of a progressively compacted material. Overconsolidated sands should however not show dilation properties, and unloading would thus have little effect. However, the amounts of silt and clay present throughout the Neogene aquifer sediments might initiate such dilation properties. Moreover, dissolution of certain mineral phases or framework grains by meteoric water might also enhance permeability, as shown by Lambert *et al.* (1997).

The objectives of this paper are therefore i) to test whether the hydraulic conductivity and its spatial heterogeneity in outcrops obtained through air permeametry are comparable to those of nearby aquifer and aquitard sediments, ii) to evaluate major differences between outcrop and aquifer sediment *K* heterogeneity including the transferability of information from outcrop to aquifer sediments, and iii) to discuss the scale effect and overall outcrop parameter representativity for use in groundwater modelling. For this purpose the results from the outcrop study by Rogiers *et al.* (2013a; Chapter 2) are compared with more standard borehole core analyses and pumping test results. Moreover, grain size analyses are used to verify the similarity between outcrop and subsurface sediments. In a final step, we provide possible explanations for the observed differences in *K* behaviour and options on how to integrate air permeametry-based data with existing knowledge available from borehole and pump test analyses in view of developing more reliable groundwater flow models.

4.2. Materials and methods

4.2.1. Hydrogeological setting and outcrop analyses

Rogiers *et al.* (2013a; Chapter 2) proposed a methodology to characterize small-scale *K* variability from outcrops, and at the same time obtain outcrop-scale equivalent *K* values. This methodology relies on air permeability measurements that are converted to saturated *K* values using the empirical equation from Iversen *et al.* (2003), and a subsequent numerical upscaling step. This methodology was tested on five outcrops from three key formations of the Neogene Aquifer in north-eastern Belgium (from top to bottom): the Mol Formation (the abbreviation Fm. will be used in the subsequent discussions), sandy and clayey parts of the Kasterlee Fm., and the clayey and sandy parts of the Diest Fm. The bias of the results from the 2D approach compared to the 3D reality is estimated to be not more than 16% for the outcrops considered here (larger *K* for 3D; see Chapter 2).

For these five formations additional geological and hydrogeological data is available from a recent characterization campaign (Beerten *et al.*, 2010) of the shallow aquifer sediments in Mol/Dessel (up to about 40 m depth), including seven cored boreholes (Figure 2.1C). This lithostratigraphical succession and its main characteristics are presented in Figure 4.1. Apart from the minimum and maximum unit thickness obtained from this recent characterization campaign, typical borehole cores are displayed for each unit, as well as a typical cone penetration test (CPT), grain size and glauconite content profile through most of the units. The

most striking features are the high clay and fine silt contents within the aquitard represented by the clayey part of the Kasterlee Fm., the sudden increase of the glauconite content in the sediments below this unit, and the contrast in coarse sand content between the upper and lower aquifers separated by the aquitard.

In addition to the individual air permeameter measurements (measurement scale of several cm^2) and their statistics, the measurement grids were numerically upscaled to obtain equivalent horizontal and vertical *K* values at the scale of the outcrop (*i.e.* typically several m^2; Rogiers *et al.*, 2013a; Chapter 2). Moreover, the air permeameter results were validated using independent constant-head laboratory permeameter measurements on 100 cm^3 ring samples taken from the same outcrop measurement grid. Therefore, the *K* data obtained from the outcrops is deemed accurate and unbiased. An overview of all outcrops characterized by air permeameter measurements within the study area is provided by Rogiers et al. (2013b; Chapter 3).

4.2.2. Constant head *K* measurements

To characterize the aquifer sediments' hydraulic conductivity variability, multiple undisturbed 100 cm^3 ring samples (with diameter of 53 mm) were taken from contiguous borehole cores (Beerten *et al.*, 2010). The ring samples were pushed in the cores in horizontal or vertical direction, for characterization of respectively horizontal or vertical *K*. The gathered data enclose several hundred hydraulic conductivity measurements on such 100 cm^3 ring samples from 7 cored boreholes, representing 350 m of core material. Two samples were taken each 2 meter, for horizontal and vertical *K*, but the anisotropy at the sample scale was generally negligible (Beerten *et al.*, 2010). The average thickness of the Mol and Kasterlee Formations in these boreholes is respectively 20 and 10 m. The highly stratified clayey part of the Kasterlee Fm. – coarse sand layers alternate with heavy clay lenses with thickness varying from less than a cm to several cm – varies in thickness from 2 to 6 m. The Diest Formation is not penetrated fully, but was characterized on average across 15 m.

All 100 cm^3 ring samples were analysed in the lab using the constant head method (Klute, 1965), using a low-pressure device for coarse material and a high-pressure device (approx. 6 bar) for the clay material expected to display low *K* values (see Beerten *et al.*, 2010 for more details). Total porosity was also determined for most core samples, as well as bulk density and volumetric moisture content for the outcrop samples, by repeatedly weighing the samples after drying and complete saturation. The methodology is similar to that used by Rogiers *et al.* (2013a; Chapter 2) to validate the outcrop air permeameter measurements.

4.2.3. Grain size measurements

A sedigraph or a combination of standard sieving and a suspension cylinder (European standard EN 933-1) was used to quantify respectively 20 and 8 grain size fractions of the borehole core samples. All samples were prepared by removing carbonates and organic matter. Clay samples were analysed with the first method, after removing particles larger than 250 μm by sieving. For more details on the data, the reader is referred to Beerten *et al.* (2010) and Rogiers *et al.* (2012; Chapter 6).

Grain size analyses of outcrop samples were performed by laser diffraction with a Malvern Mastersizer (Malvern Instruments Ltd., UK). This method consists of monitoring the amount of reflection and diffraction that is transmitted back from a laser beam directed at the particles, and quantifies 64 grain size fractions. Each sample was divided into 10 sub-samples by a rotary sample splitter to enable repeated measurements on a single sample, and all samples were measured at least twice. The final result was based on the average grain size distribution of all sub-samples. Note that particle sizes are expressed as the size of an equivalent sphere with an identical diffraction pattern.

4.2.4. Pumping tests

Step drawdown, constant discharge and recovery tests were performed at different locations within the study area, including some of the borehole locations. The transient groundwater head observations were interpreted with analytical as well as numerical models (Meyus and Helsen, 2012). Results from these large-scale tests are used here to illustrate the scale effect for hydraulic conductivity determination on subsurface sediments, and to compare such large-scale measurements with

the numerically upscaled *K* values for the outcrops.

4.2.5. Variography

The experimental variograms are all fitted with spherical models, using a weighted least squares approach. Two approaches are tested: 1) treating outcrop and borehole data sets separately (variogram models for the outcrops are taken from Rogiers *et al.*, 2013a; Chapter 2), and 2) using a pooled data set which combines both outcrop and borehole data. In the latter case equal weight is given to both datasets in the least squares fitting. In the former case individual experimental variogram points are

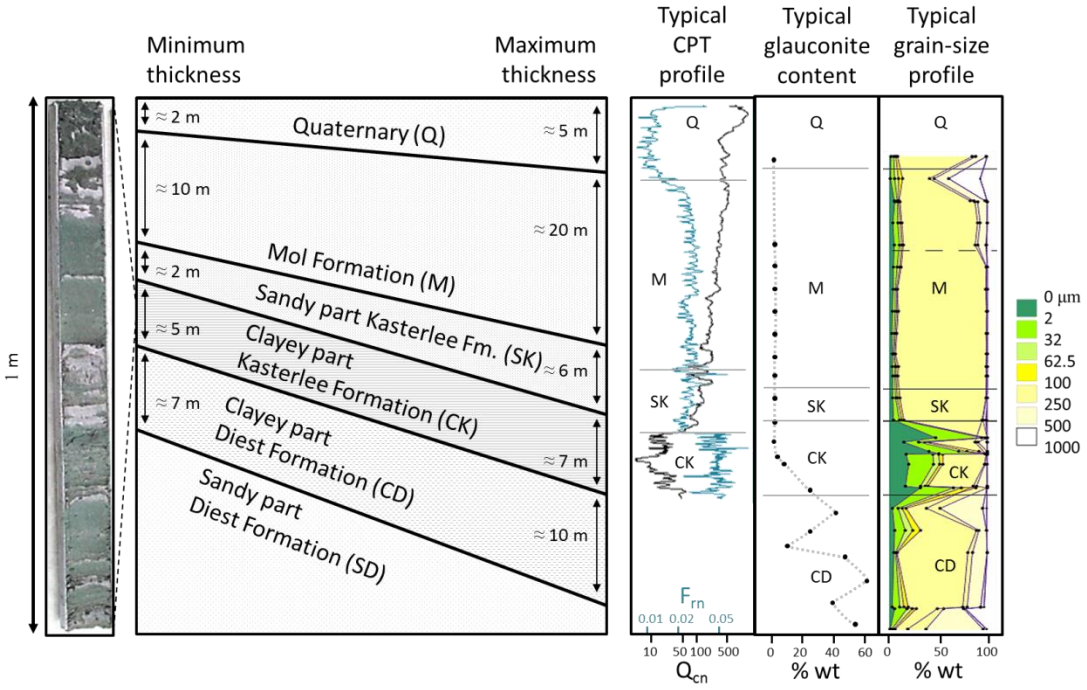

Figure 4.1: Overview of the studied lithostratigraphical succession with formation thicknesses, a typical CPT profile (Q_{cn} and F_{rn} are the normalized cone resistance and friction ratios), typical glauconite content (weigth percentage; % wt), and a typical grain-size profile. A picture of a borehole core from the clayey part of the Kasterlee Formation is provided to illustrate its heterogeneity. For more information, see Beerten *et al.* (2010).

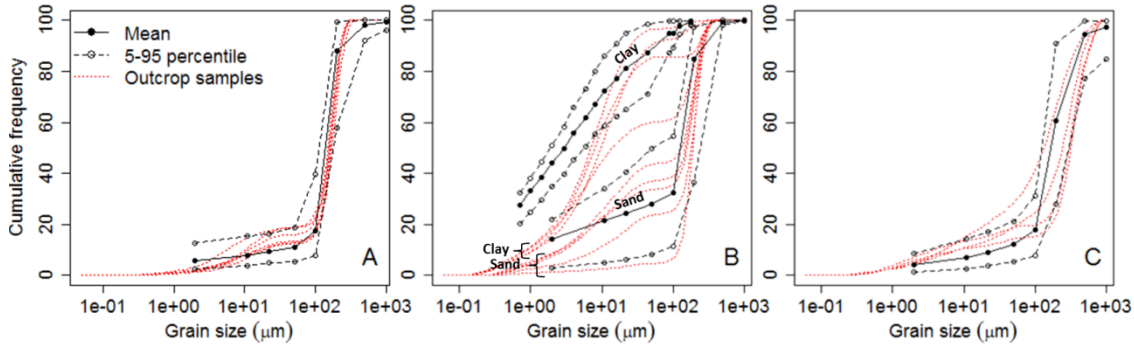

Figure 4.2: Cumulative grain size distributions for the outcrop (laser diffraction) and borehole data (mean value and 5-95 percentiles from sedigraph or standard method; Beerten *et al.* 2010) for A) the sandy Kasterlee Fm., B) the clayey Kasterlee Fm. and C) clayey Diest Fm.

weighted according to the number of point pairs they represent. The initial variogram parameters for the nugget, total sill and range were respectivley set to the overall minimum semivariance, the data variance, and the maximum lag distance. In certain cases singular model fits occurred due to non-uniqueness (data does not allow to discriminate between different equivalent models, *e.g.* pure nugget vs spherical model with zero range). The responsible parameters were then fixed at their initial value, before re-initialising the model fitting procedures. All variography was performed with the gstat package (Pebesma, 2004).

4.3. Results and discussion

4.3.1. Grain size distributions

Prior to comparing *K* values obtained from different measurement methods, a comparison is made between grain size distributions for the outcrop sediments and aquifer materials collected from cored boreholes (Beerten *et al.*, 2010). This evaluation is necessary to verify if the outcrop and aquifer sediments represent the same lithostratigraphical units, and to highlight possible discrepancies between both to inform the comparison of their corresponding *K* values.

Overall there is good correspondence between outcrop/aquifer grain size distributions for the sandy part of the Kasterlee Fm. and clayey part of the Diest Fm. (Figures 4.2A-C), with a somewhat larger fraction of fines (*i.e.* between 2 and 22 μm) for the outcrop samples. Van Ranst and De Coninck (1983) suggested that post-depositional weathering of glauconite material, a green iron-rich clay mineral, might increase the relative amount of fines. Kasterlee Formation samples collected from boreholes contain glauconite up to a few percent, but for the Diest Formation it is at least 10 to 20% (Beerten *et al.*, 2010). The disintegration of the glauconite fractions in the outcrops could thus have increased the fines content.

The comparison further illustrates that the clay fraction (< 2 μm) of the clayey part of the Kasterlee Fm. is about 20% lower in the outcrop samples compared to the aquifer material. Since we are dealing with outcrop samples that are close to the surface, post-depositional migration of clay out of the clay lenses (*e.g.* Mažvila *et al.*, 2008) together with bioturbation in the outcrops is a plausible explanation for the lower clay content in the outcrop. For the clayey Kasterlee Fm. outcrop, the individual grain size distribution curves (Figure 4.2B) indicate a continuous gradation between two extreme cases, *i.e.* from a clay lens texture (approximately 40% clay) to coarse sand without fines (>90% sand). The corresponding grain size distributions for boreholes show no overlap between the clay and sand samples, an illustration of the existence of two distinctly different materials within the clayey part of the Kasterlee Fm., *i.e.* heavy clay lenses embedded in coarse sands characterized by a sharp interface (Beerten *et al.*, 2010).

In conclusion, weathering, clay migration, and bioturbation may have influenced the lower end of the outcrop samples' grain size distribution considerably. Furthermore, dissimilarities in palaeogeographic conditions and sediment source regions between the outcrop and borehole locations may equally explain such differences. However, the consistent stratigraphic position of the clayey Kasterlee Fm. sediments on top of the Diest Fm. and the relatively good correspondence in particle size for the sandy material (*i.e.* sand layers within the Kasterlee Fm.), are sufficient underpinning arguments to support using the studied clayey Kasterlee Fm. outcrop at Heist-op-den-berg (for details of the outcrop see Rogiers *et al.*, 2013a; Chapter 2) as surrogate for the clayey Kasterlee Fm. aquitard (Gulinck, 1963, Laga, 1973; Fobe, 1995). Additional insight could be obtained from tracing the exact origin and initial composition of the outcrop materials; however, this is beyond the scope of the current paper.

4.3.2. Hydraulic conductivity distributions

Figure 4.3 provides a comparison of outcrop and borehole (aquifer) *K* kernel density estimates of the probability density functions (pdfs) for the five sediments. Statistically significant differences exist for all sediments, with *p* values for *F*-tests all below 4×10^{-3}, while the corresponding *t*-tests *p* values are all below 1×10^{-5} indicating statistically significant differences for both the variance and mean. All outcrop pdfs have higher mean *K* values than their borehole complement. While most outcrop samples display conductivities between 10^{-5} and 10^{-3} m/s, borehole samples have their most frequent *K* values between 10^{-6} and 10^{-4} m/s. Moreover, the standard deviations for the borehole samples are consistently larger than

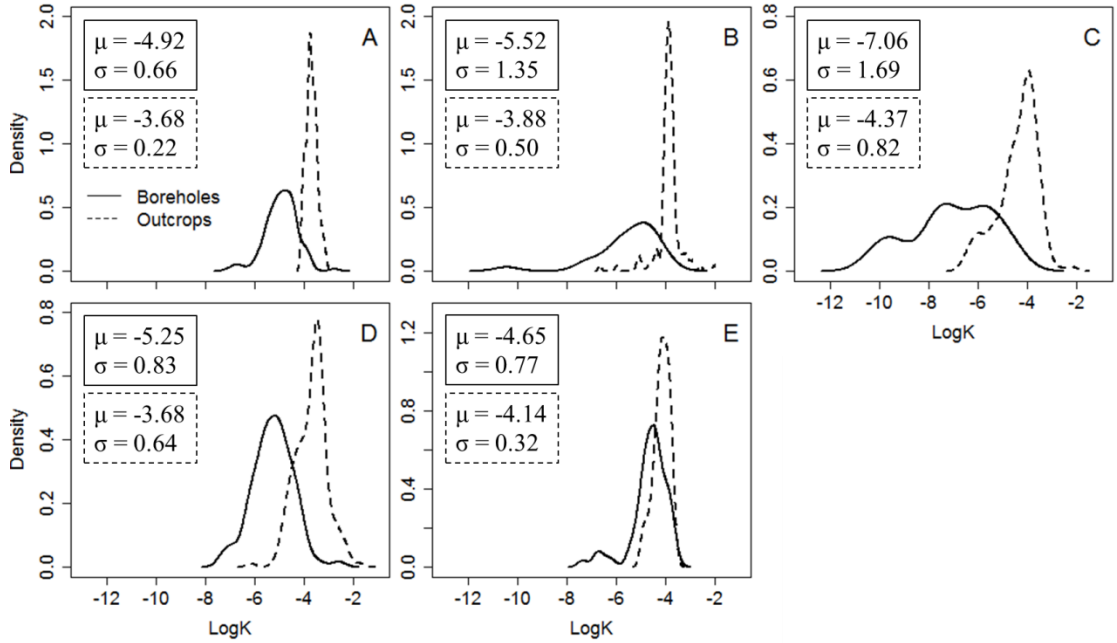

Figure 4.3: Comparison between distributions (kernel density estimates of the probability density functions) for air permeameter based outcrop *K* and constant-head *K* measurements on undisturbed samples from cored boreholes, for A) the Mol Fm., B) the sandy Kasterlee Fm., C) the clayey Kasterlee Fm., D) the clayey Diest Fm. and E) the sandy Diest Fm. mean (μ) and standard deviation (σ) are given for both data sources.

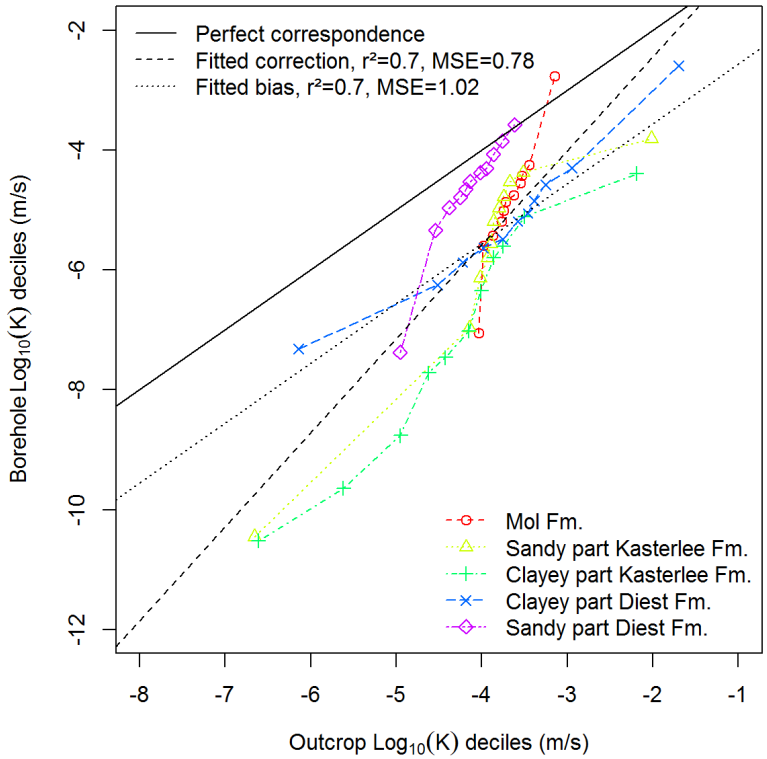

Figure 4.4: Outcrops versus borehole $\log_{10}(K)$ deciles, a fitted linear correction model (y = 0.6938 + 1.5685 x), and a fitted bias (y = x – 1.559).

those based on the outcrop samples. The left tail of the pdfs tends to be much larger for the borehole data while the peaks tend to be wider (one to two orders of magnitude for the outcrops versus two to four orders of magnitude for the borehole data), especially for the sandy Kasterlee Fm. (Figure 4.3B). Relative variability expressed as coefficient of variation (*CV)* is approximately two times larger for borehole pdfs than for outcrop pdfs (Mol Fm.: -13.4% versus -5.9%; Kasterlee Fm. sands: -24.5% versus -12.9%; Diest Fm. sands: -23.9% versus -18.8%) while it is similar for the clayey parts of the Kasterlee Fm. (-23.9% versus -18.8%) and Diest Fm. (-15.8% versus -17.4%). For the borehole data, sampling occurred over a large geographical area (several tens of km^2 versus as little as a few m^2 to at most a few tens of m^2 for the outcrops) and over a much larger depth (up to 50 m) thus having the opportunity to sample a much larger spatial heterogeneity. Several characteristics typical of heterogeneity in *K* are however visible in both the outcrop and borehole *K* distributions.

For the sandy part of the Kasterlee Fm. (Figure 4.3B), a long tail towards low values is present both in the outcrop and in the boreholes, while the majority of samples is within a much narrower distribution in the outcrop. For the clayey part of the Kasterlee Fm. (Figure 4.3C), a multi-modal distribution is present for both datasets and representative of samples belonging mainly to clay lenses or sand layers. The clayey part of the Diest Fm. (Figure 4.3D) displays a similar pdf in both datasets (ratio of borehole to outcrop CV = 0.91), and the sandy Diest Fm. data (Figure 4.3E) shows the best absolute match in terms of the mean *K*, although the second peak with lower *K* values was not observed in the outcrop.

Validation of air permeameter *K* with core-based outcrop *K* demonstrated absence of systematic bias in the air permeameter *K* estimates (Rogiers *et al.*, 2013a; Chapter 2). Therefore, differences in *K* distributions between outcrop and aquifer sediments can be attributed to the scale of investigation (a single outcrop with a typical measurement grid of a few m^2 versus seven ~50-m deep vertical transects through the different lithostratigraphical formations, Figure 4.1), different evolutionary states of the outcropping and subsurface sediments, and possibly different sedimentation conditions.

4.3.3. Linear rescaling correction and bias

To investigate the (dis)similarities between the outcrop and borehole data across these five lithological units, the minimum and maximum values are plotted in Figure 4.4, with all deciles (10^{th}, 20^{th}, ..., 90^{th} percentile) in between. This shows that linear scaling of the outcrop values to the corresponding borehole distributions is possible for all outcrops. The extreme values are however not always in line with the centre of the distributions (as indicated by the deviation of the overall shape of the first and last line segments). All outcrops exhibit a more or less similar trend for at least part of the data, which is supported by the linear model fit on all minimum, maximum and decile points (r^2 = 0.7). The slope, larger than 45°, indicates that the deviation between outcrop and boreholes is larger for low *K* than for higher *K* values, which is consistent with the previous observations. The sandy Diest Fm. curve lies apart and above the other curves, and is much closer to the 1:1 line of perfect agreement. This is as expected based on the good correspondence in pdfs (see Figure 4.3E). In other words, the Diest Fm. outcrop is well and truly representative for the entire aquifer unit.

This linear rescaling however accounts for the discrepancy in representativity between the outcrop and borehole campaigns. The variance from the outcrops is adjusted, to estimate the more representative borehole *K* distribution. However, when comparing *e.g.* the spatial variability from both datasets, we do not want this discrepancy to influence the results. Therefore, we make use of a fitted bias (or linear model with an intercept, and slope 1; Figure 4.4) for that purpose, which does not alter the spatial variability.

4.3.4. Porosity and compaction state

Weathering of clay layers at the surface has certainly contributed to produce higher *K* values for the fine material in the outcrops, but the systematic bias of about one order of magnitude that is also present for the sands remains unexplained.

Trends in porosity or bulk density with depth are very hard to detect in the borehole data due to the extensive layering of different lithologies and grain size distributions at the study area (the same lithology may occur at different depth

depending on the geographical location). Moreover, the data from the outcrops are hardly sufficient to prove differences with the subsurface sediments are statistically significant. For example, the mean total porosity for the four Mol and Kasterlee Fm. outcrop core samples is 43% with a mean dry bulk density of 1.52 g/cm^3 (see Rogiers *et al.*, 2013a; Chapter 2), while the borehole values of the same two formations (43 samples) are 40% and 1.60 g/cm^3 (samples between 2 and 28 m below surface). This is consistent with different compaction states, *i.e.* outcrop samples being less compacted than borehole samples, but the differences remain very small and are only significant for porosity at the 5% significance level. However, even small differences in porosity can yield large differences in *K* (see discussion below).

The impact of the degree of compaction on *K* values was further investigated for the borehole dataset only using total porosity as proxy for compaction, as analyses in literature show that porosity has a high influence on *K*, given a homogeneous grain size distribution and chemistry (*e.g.* Bourbie and Zinszner, 1985). On an individual sample basis, it is hard to detect total porosity – *K* relationships within the borehole dataset, since these are very complex owing to the influence of grain size (Rogiers *et al.*, 2012; Chapter 6), sorting, packing and eventually the actual accessible pore throat radii (*e.g.* Bakke and Øren, 1997; Øren *et al.*, 1998). However, as indicated by the scatterplot in Figure 4.5, if total porosity and *K* are averaged for each formation and for each borehole separately, some statistically significant relationships exist. The slopes of the linear model fits are consistently positive, and in several cases, a change of a few percent in porosity can change *K* drastically. For instance, a one % decrease in porosity yields a decrease in *K* of minimum 0.14 and maximum 1.08 $\log_{10}$ units. This is a partial confirmation of the importance of the degree of consolidation and compaction on our *K* values; corroborating evidence about the effect of grain size, sorting and packing characteristics will be sought in future research.

An additional analysis of the *K* – depth below surface relationship was performed but did not yield any significant dependencies (results not shown). This is probably due to the alternation of different lithologies and grain sizes with depth, hence obscuring the influence of depth on compaction and thus on porosity and *K*.

4.3.5. The scale effect and vertical anisotropy

The representativity of *K* measurements – whether for outcrop or aquifer sediments – for characterizing a lithostratigraphical unit depends, among others, on the size of the measurement scale (or measurement support) and the spatial extent and lithostratigraphic complexity of the sampled domain. The effect of measurement scale for individual *K* measurements also impacts the overall variability, as measurements with a larger support volume, like pumping tests, average out the small-scale variabilities (Mallants *et al.*, 1997). It is thus important in the comparison between outcrop and borehole *K* values to consider such scale-effects.

A comparison between the outcrop data (air permeameter based geometric mean *K* values and the calculated corresponding equivalent *K* values) and the subsurface data (borehole core geometric mean *K* values and the pump test *K* values) is shown in Figure 4.6. It reveals the overall range is smallest for the outcrop data, both at the smallest measurement scale (data for air permeameter measurements spans 5 orders of magnitude versus 8 orders of magnitude for borehole cores) and at the largest scale (calculated equivalent outcrop *K* values show a range of ~2 orders of magnitude versus ~5 orders of magnitude for pump tests). It is further evident that the outcrop-based equivalent *K* values are systematically higher than the mean borehole core values; a better correspondence is achieved with the pump test values.

Because a pump test represents a large support volume, easily tens to hundreds of m^3, small-scale heterogeneities have much less effect on such large-scale *K* values, hence the smaller data range. Furthermore, the support volume is commensurate with the computational domains used to calculate equivalent outcrop values. Overall the pump test values are generally only slightly smaller than the equivalent outcrop values, except for the clayey part of the Kasterlee Fm. for which the discrepancy is about three to four orders of magnitude. This again emphasizes the need for a correction if outcrop *K* values are used to inform building

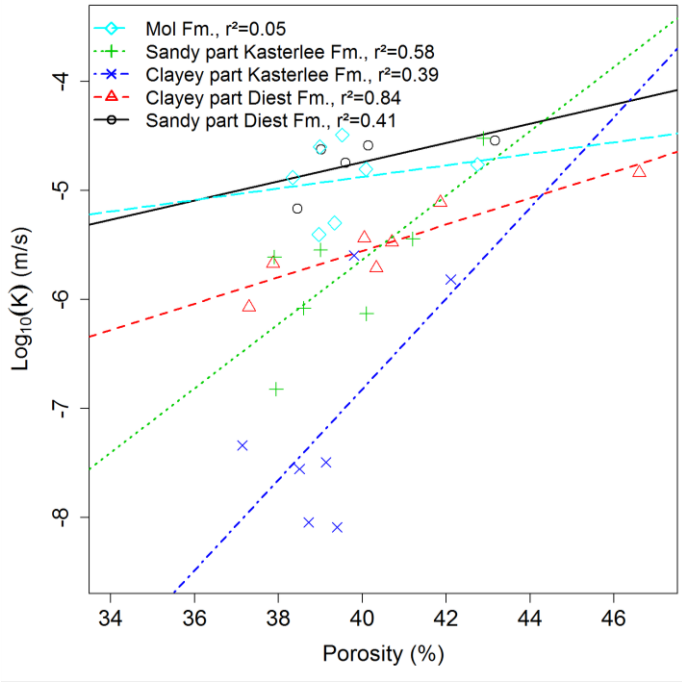

Figure 4.5: Scatterplot of $\log_{10}$-transformed hydraulic conductivity *K* versus porosity (borehole dataset only) for the five lithostratigraphical units with corresponding linear model fits. Each data point represents the mean porosity and mean *K* of all measurements pertaining to one formation for one particular borehole.

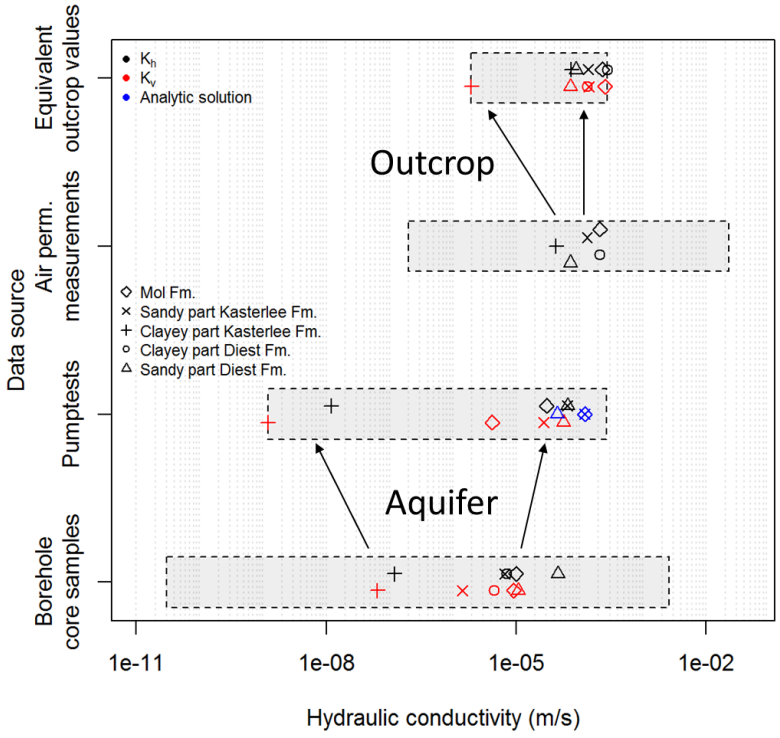

Figure 4.6: Comparison of geometric mean *K* values obtained from borehole core samples, pump tests, outcrop air permeameter measurements and calculated equivalent values. The gray boxes represent the data limits, and the arrows indicate the contrasting effects of upscaling for the aquifer and aquitard units.

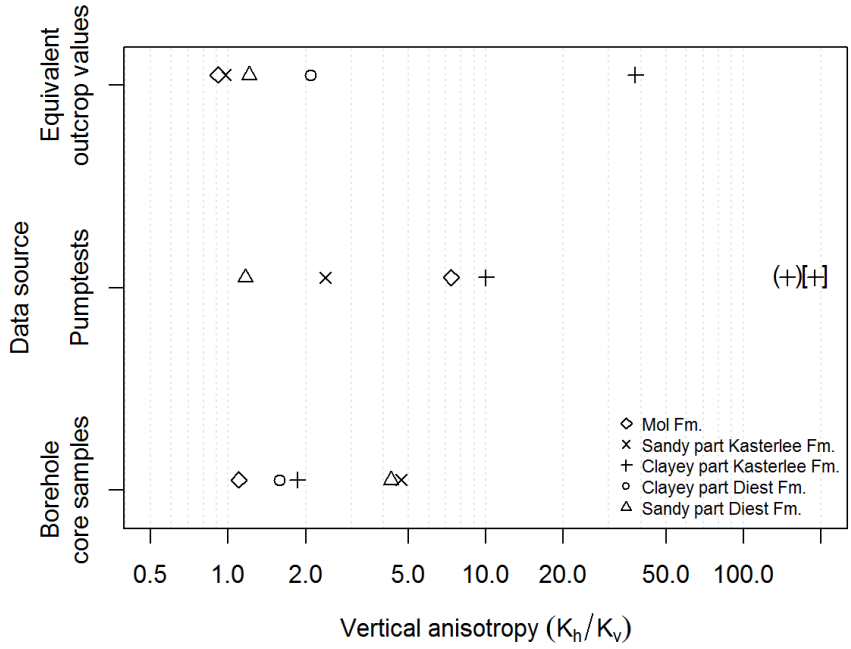

Figure 4.7: Comparison of the vertical anisotropy factors derived from the geometric mean *K* values from Figure 4.6. The plusses between round and square brackets represent respectively the parameter value obtained by Gedeon and Mallants (2012) using regional inverse modelling (with fixed K_h value) and the value representing a part of the aquitard in the original Dessel 2 pump test interpretation by Lebbe (2002) (based on borehole cores).

conceptual groundwater models. Correction models such as those from Figure 4.4 would account for impacts of different compaction and/or weathering processes, especially for the more clay-bearing sediments. The bias introduced by the 2D – 3D discrepancy for the equivalent outcrop *K* values (about 16% max) seems to be negligible compared to the orders of magnitude on which the other values are spread.

The arrows in Figure 4.6 indicate different effects of upscaling for the aquifer and aquitard units. Moving from the sample (cm-scale) to the pumptest-scale (meter-scale) in most cases increases the aquifer geometric mean *K* values by one order of magnitude, while the outcrop values remain more or less constant when geometric means are compared with effective values. Unlike the other formations, upscaling the clayey part of the Kasterlee Fm. data results in a decrease of the average *K* values, for both K_v and K_h pertaining to the aquifer and for outcrop K_v. This indicates that in both the outcrop and aquifer sediments of this particular lithostratigraphic unit a significant amount of small-scale heterogeneity is present (*i.e.* clay lenses) which significantly decreases the magnitude of the calculated effective *K* values.

Faulting could be another process involved enhancing discrepancies between small and large measurement supports. However, this process is considered to be absent as the study area is known as a zone of low seismic and limited tectonic activity (De Craen *et al.*, 2012).

A comparison of the vertical anisotropy values (K_h/K_v) is shown in Figure 4.7. The K_h/K_v ratios based on the geometric means of the 100 cm³ borehole cores lies between 1 and 5. The two lithostratigraphical units with the highest K_h/K_v values are the sandy parts of the Kasterlee and Diest Fm., which are influenced by some outliers (see Figure 5.3). The equivalent outcrop K_h/K_v values are less than the corresponding borehole core anisotropy values, except for the clayey parts of the Diest and Kasterlee Fm. For the latter K_h/K_v increased more than one order of magnitude, when moving from the borehole core to the outcrop scale. The pump test anisotropy values mostly show larger values compared to those from the borehole cores, with a maximum vertical anisotropy of 10. The original Dessel 2 pump test interpretation by Lebbe (2002) yielded *K* values for the clayey part of the Kasterlee Fm. and mentions a vertical anisotropy factor of 190 for part of the aquitard.

This value was obtained by inverse modelling of the pump test, but due to a limited drawdown across the aquitard, the optimized parameter values remain highly uncertain. A more reliable estimate was obtained from the regional modelling of the Neogene aquifer and the flow across the aquitard by Gedeon and Mallants (2012). They obtain a vertical anisotropy of 148 by inverse conditioning on regional piezometric observations above and below the aquitard. The high vertical anisotropy determined from the outcrop (K_h/K_v = 38) supports these values, and indicates that such large values might be more realistic at larger scales.

4.3.6. Spatial variability

The vertical spatial variability for the outcrop and borehole data (K_h only) is compared in Figure 4.8 and Table 4.1. For the Mol Fm., the outcrop data overall shows less variability (smaller semi-variance) than the borehole core samples; but correspond well with the experimental borehole variogram at the centimeter to meter scale. The larger total sill for borehole (0.13 + 0.41) compared to outcrop (0.05) is a reflection of the larger variability captured by the borehole data. This larger variability is caused in part by combining two local stratigraphical sub units into the Mol Fm. (see Beerten *et al.*, 2010) with thin gravel layers and clay lenses at their interface. The borehole data also displays a larger vertical spatial range (*i.e.* 20 m) than the outcrop (*i.e.* pure nugget), owing to samples being collected from a much larger vertical sampling window (up to 20 m) and multiple boreholes spread over several km^2. As both experimental semivariograms are compatible, fitting the joint dataset improves the variogram model fit considerably.

For the Kasterlee Fm. sands, the borehole and outcrop data show a large difference, which might be due to the rather limited number of borehole core samples identified as the sandy part of the Kasterlee Fm. or an increased amount of heterogeneity in the outcrop due to weathering processes. The overall variability (total sill) is more or less similar for both outcrop and borehole data, suggesting that the variability captured by the outcrop samples may be used as surrogate for the variability in boreholes. Despite the presence of spatial correlation in the both datasets, the joint model fit shows a pure nugget because of the high semivariance values for the outcrop data.

The clayey Kasterlee Fm. shows the largest spatial variability of all lithological units for both the outcrop and borehole data. While the outcrop shows some spatial correlation, the borehole model shows a pure nugget. The borehole cores show higher variability due to the clay-rich lenses and correspondingly low *K* values, which are altered in the outcrops, but only the first data point at 0.5 m is contradicting the outcrop data. The joint model fit (Figure 4.8) does reveal their compatibility, and shows spatial correlation up to a few meters. This model might be more useful than the individual variogram models due to the integration of different scales.

Most of the clayey Diest Fm. outcrop data seems to be compatible with the borehole core spatial variability. All three model fits show a range of one to two meters, and similar total sills. The sandy Diest Fm. also exhibits similar total sill in all three cases, with a larger spatial range for the borehole data. The joint model fit is compatible with that of the borehole data, but shows a higher nugget due to the higher semivariances in the outcrop data.

Overall, the borehole data exhibit larger correlation lengths than the outcrop data. The total sills are mostly similar, except for two cases were the borehole data clearly encompasses more heterogeneity. Three out of five experimental variograms are overlapping at certain locations, indicating that at certain scales both datasets exhibit similar spatial variability. Fitting of the joint datasets results in these cases in more robust variogram models. This indicates that small-scale structural information, such as alternation of relatively thin clay and sand layers, and its effect on spatial variability in *K* may be preserved in outcrop sediments. Therefore, analysis of outcrop stratigraphy and hydraulic conductivity variability can yield valuable qualitative and quantitative insight about such properties for similar aquifer and aquitard sediments.

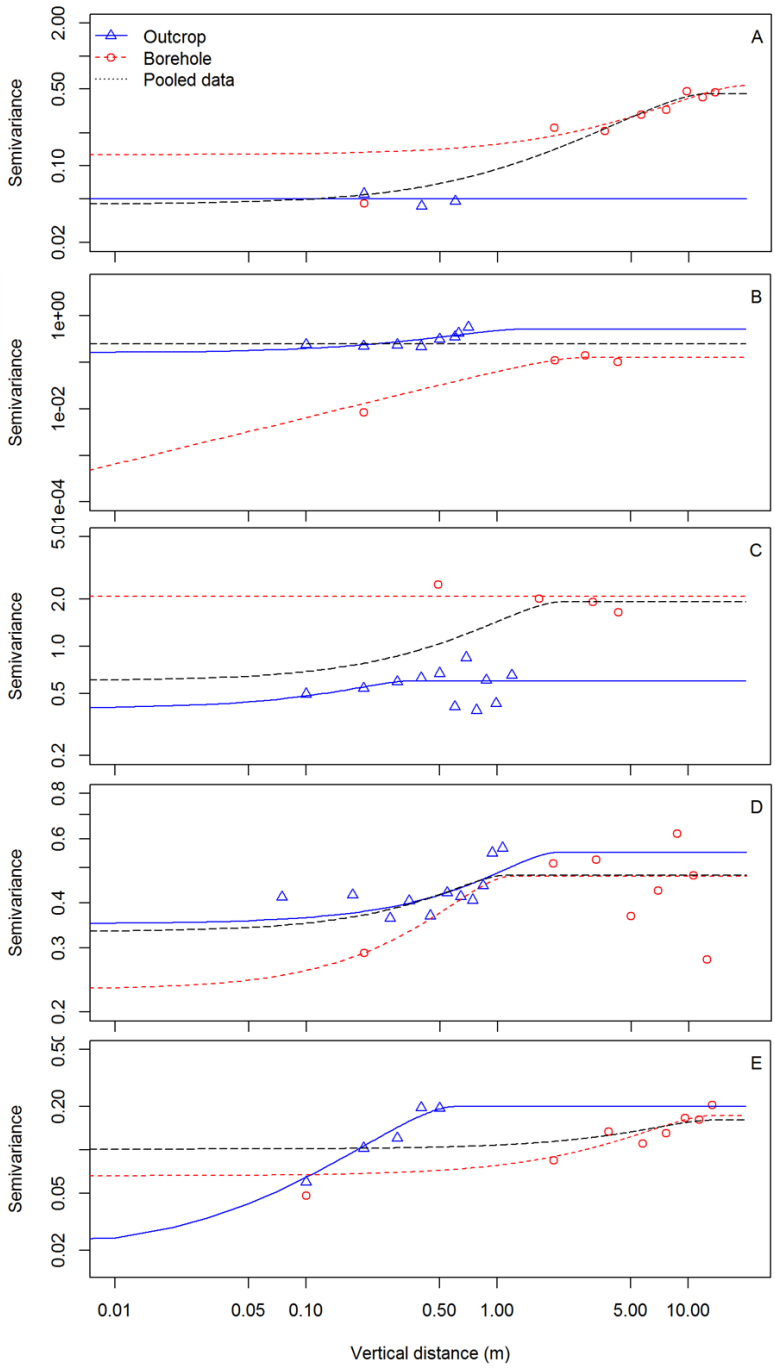

Figure 4.8: Comparison between vertical experimental and modelled semivariograms (fitted using a least squares approach) for outcrop and borehole data. A) Mol Fm., B) sandy Kasterlee Fm., C) clayey Kasterlee Fm., D) clayey Diest Fm., and E) sandy Diest Fm.

Table 4.1: Overview of fitted variogram model parameters for the vertical experimental variograms (range = correlation length). The outcrop data is taken from Rogiers *et al.* (2013a (Chapter 2)).

Sediment	Parameter	Outcrop	Borehole	Pooled
Mol Formation	Nugget	0.05	0.13	0.04
	Sill	-	0.41	0.41
	Range (m)	-	19.66	12.46
	Type	---------- Spherical ----------		
Kasterlee Formation: sandy part	Nugget	0.16	0.00	0.25
	Sill	0.35	0.13	-
	Range (m)	1.36	2.90	-
	Type	---------- Spherical ----------		
Kasterlee Formation: clayey part	Nugget	0.40	2.07	0.60
	Sill	0.20	-	1.32
	Range (m)	0.36	-	2.20
	Type	---------- Spherical ----------		
Diest Formation: clayey part	Nugget	0.35	0.23*	0.33
	Sill	0.20	0.24	0.14
	Range (m)	2.07	1.17	1.12
	Type	---------- Spherical ----------		
Diest Formation: sandy part	Nugget	0.02	0.07	0.10
	Sill	0.18	0.11	0.06
	Range (m)	0.60	13.34*	13.34*
	Type	---------- Spherical ----------		

*fixed during variogram model fit

4.4. Perspectives

Despite the limitations of and systematic differences between the outcrop and borehole datasets, we have demonstrated that outcrop studies can provide useful information for developing more reliable groundwater flow and contaminant transport models. Because of the systematic differences observed here between outcrop and subsurface sediments, the obtained outcrop *K* values are not directly applicable in groundwater flow modelling, unless a correction is applied (Figure 4.4). Furthermore, the different *K* distributions are comparable at least in a relative way, and linear scaling based on deciles was shown to be relatively accurate. In other words, results such as the spatial heterogeneity models, the equivalent vertical anisotropy factors, and relative differences between the different sediments provide us with information useful to guide conceptual groundwater flow model building and constraining model parameterization.

Potential applications of our findings for building conceptual and numerical models of groundwater flow include i) where possible highly structured heterogeneity should either be represented explicitly in the models or use should be made of appropriate geostatistical tools (*e.g.* multiple point statistics) based on detailed structural information visible in and quantifiable from outcrops, ii) use of the obtained equivalent vertical anisotropy factors can influence conceptual model choices for isotropy/anisotropy for certain units, and the actual value represents a minimum of the parameter range in larger scale groundwater flow simulations (especially in a layered stratigraphical setting), iii) to avoid over-parameterization, ratios between *K* values of different units can be fixed during model optimization (*e.g.* Gedeon and Mallants, 2012) using the ratios obtained from equivalent outcrop estimates, and iv) use of the obtained outcrop variogram models can complement information from a larger scale (*e.g.* boreholes),

or be used for small-scale geostatistical simulations for detailed local transport simulations. All these applications will be most beneficial when combined with the traditional borehole coring and measurements and other invasive and non-invasive subsurface characterization techniques.

4.5. Conclusions

Analysis of outcrop sediments considered to be analogues for various lithostratigraphical units within a sedimentary aquifer provided a qualitative understanding of aquifer and aquitard stratigraphy and a quantitative estimate about *K* variability at the centimeter- to meter-scale. Comparison between outcrop and independent borehole core *K* values revealed significant differences between both datasets. Such differences are believed to be induced mainly by weathering, different palaeoenvironmental conditions and differential compaction, and can be corrected for as was demonstrated on the basis of a linear model. Hence, outcrop information can be used for building better stratigraphic models including determination of spatial structure by variogram fitting for further use in geostatistical simulations. Moreover, the relative variability in *K* values with similar coefficients of variation for borehole and outcrop *K*, and the derived anisotropy values are very useful to get a more complete understanding of the heterogeneity within the Neogene Aquifer.

Comparison of outcrop and borehole *K* values demonstrated the borehole *K* probability density functions had broader peaks, longer tails towards low values, and the presence of a systematic bias. The reasons behind this discrepancy are manifold, and include weathering of the outcrop sediments and a lesser degree of consolidation and associated stress states in outcrops. Also, measurements performed on outcrops sometimes several tens of kilometres away from the main study site may further invoke differences in *K*. Grain-size analyses showed that the sediments from the investigated outcrops and boreholes are similar but not necessarily exactly the same. Sorting might play a role as well. Clay migration and bioturbation in the outcrop sediments probably contributed to the observed discrepancies, as well as slight differences in palaeoenvironmental settings. The degree of (over)consolidation and stress states might also have an impact, mainly through porosity, but further research is needed to confirm or quantify this, as trends with the current depth of the sediments are hard to detect due to the alternation of different lithologies.

Based on all data a linear scaling relationship was derived (r^2 = 0.7) that permits rescaling of outcrop *K* distributions to their subsurface equivalents. For most individual units, the differences between outcrop and subsurface sediments were similar (except for the extremes of the distributions). The sandy part of the Diest Fm. however showed a considerably better fit between outcrop and aquifer than the other cases.

In a comparison with *K* values obtained through other means, outcrop-based equivalent *K* values were systematically higher than those from pump tests (especially for the clayey part of the Kasterlee Fm.), whose support volumes are considerably larger than the simulation domains considered in the outcrops. Mean borehole core samples resulted in the overall smallest *K* values. Smaller compaction at shallow depth and long-term biophysical weathering processes presumably contributed to outcrop equivalent *K* values being larger than any other estimate of large-scale *K* available in this study.

In most cases the semivariograms for the outcrop and borehole data are compatible. Only for the sandy Kasterlee Fm. the outcrop data clearly shows higher variability than the borehole data. Spatial correlation (*i.e.* increasing semivariance with distance) is present in most cases, either in the outcrop or borehole data, or both. The clayey Diest Fm. shows however a pure nugget effect for both datasets. For the Mol Fm. and the clayey Kasterlee Fm. both datasets complement each other resulting in more robust semivariogram model fits. For the sandy Diest Fm. there seems to be a discrepancy in the range between both datasets.

Given the small number and limited size of the studied outcrops, transfer of information from outcrops to the corresponding aquifer sediments can be improved by expanding the number of outcrops for the same lithostratigraphical units. In addition, more complementary aquifer information could be collected for developing a depth-dependency in aquifer *K* that incorporates effects of compaction which could then be used to rescale outcrop *K* values to sediment values at a given depth. Such information, together with

geostatistical parameters, may be used as input or prior information to stochastic flow models.
Next to the quantitative information tested in this paper, information about facies geometry, like the alternating clay and sand layers within the clayey Kasterlee Fm., cannot be revealed easily using available *in situ* methods, and represents very important qualitative knowledge obtained from outcrops.

References

Bakke, S., Øren, P., 1997. 3-D pore-scale modelling of sandstones and flow simulations in the pore networks. *SPE Journal* **2**: 136–149.

Bayer, P., Huggenberger, P., Renard, P., Comunian, A., 2011. Three-dimensional high resolution fluvio-glacial aquifer analog: Part1: Field study. *Journal of Hydrology* **405**(1-2): 1-9.

Beerten, K., Wemaere, I., Gedeon, M., Labat, S., Rogiers, B., Mallants, D., Salah, S., Leterme, B., 2010. Geological, hydrogeological and hydrological data for the Dessel disposal site. Project near surface disposal of category A waste at Dessel – Version 1.- Brussels, Belgium: NIRAS/ONDRAF, 2010.- 273 p.- NIROND–TR 2009–05 E V1.

Beerten, K., Deforce, K., Mallants, D., 2012. Landscape evolution and changes in soil hydraulic properties at the decadal, centennial and millennial scale: A case study from the Campine area, northern Belgium. *Catena* **95**: 73-84.

Bourbie, T., Zinszner, B., 1985. Hydraulic and acoustic properties as a function of porosity in Fontainebleau sandstone. *Journal of Geophysical Research* **90**(4): 524–532.

De Craen, M., Beerten, K., Honty, M., Gedeon, M., 2012. Geo-scientific evidence to support the I2 isolation function (geology and long-term evolution) as part of the Safety and Feasibility Case 1 (SFC1). Mol, Belgium: SCK•CEN, 2012.- 120 p.- (External Report of the Belgian Nuclear Research Centre; ER-184; CCHO 2009-0940000 Research Plan Geosynthesis). ISSN 1782-2335.

Flint, S.S., Bryant I.D. (eds)., 1993. The Geological Modelling of Hydrocarbon Reservoirs and Outcrop Analogues, Blackwell Publishing Ltd., Oxford, UK.

Fobe, B., 1995. Litologie en litostratigrafie van de Formatie van Kasterlee (Plioceen van de Kempen). *Natuurwetenschappelijk tijdschrift* **75**(2): 35-45.

Gedeon, M., Mallants, D., 2012. Sensitivity analysis of a combined groundwater flow and transport model using local-grid refinement: A case study. *Mathematical Geosciences* **44**(7): 881-899.

Gulinck, M., 1963. Essai d'une carte géologique de la Campine. Etat de nos connaissances sur la nature des terrains néogènes recoupés par sondages. Symposium Stratigraphie Néogène Nordique (Gand 1961). *Mém. Soc. Belge Géol.* **6**: 30-39.

Iversen, B. V., Moldrup, P., Schjonning, P., Jacobsen, O. H., 2003. Field Application of a Portable Air Permeameter to Characterize Spatial Variability in Air and Water Permeability. *Vadose Zone Journal* **2**(4): 618-626.

Klute, A., 1965. Laboratory measurements of hydraulic conductivity of saturated soil. In: Methods of soil analysis. Part 1. Black CA. (ed.). *Agronomy* **9**: 210-220.

Konikow, L.F., 2011. The secret to successful solute-transport modeling. *Ground water* **49**(2): 144-59.

Laga, P., 1973. The Neogene deposits of Belgium. Guide book for the field meeting of the Geologists' Association London. Belgian Geological Survey: Brussel, 29.

Lambert, M. R., Cole, R. D., Mozley, P.S., 1997. Controls on permeability heterogeneity in the Tocito Sandstone (Upper Cretaceous), northwest New Mexico. In: Mesozoic geology and paleontology of the Four Corners Region. Anderson OJ, Kues BS, Lucas SG (eds). Guidebook, 48th Field Conference. New Mexico Geological Society, 217-228.

Lebbe, L., 2002. Interpretatie van een pompproef uitgevoerd te Mol. Rapport GROMO2002/07 – RUG – Vakgroep Geologie en Bodemkunde. September 2002.

Mallants, D., Mohanty, B. P., Vervoort, A., Feyen, J., 1997. Spatial analysis of saturated hydraulic conductivity of a macroporous soil. *Soil Technology* **10**:115-131.

Mažvila, J., Vaicys, M., Beniušis, R., 2008. Causes and consequences of the vertical migration of fine soil fractions. *Žemes Ukio Mokslai* **15**(3): 36–41.

McKinley, J. M., Lloyd, C. D., Ruffell, A. H., 2004. Use of Variography in Permeability Characterization of Visually Homogeneous Sandstone Reservoirs With Examples From Outcrop Studies. *Mathematical Geology* **36**(7): 761-779.

Meyus, Y., Helsen, S., 2012. Multi en enkelvoudige pomptests in en rond de nucleaire zone van Mol-Dessel (terreinverkenningen 2008), NIROND-TR 2012-19.

Øren, P., Bakke, S., Arntzen, O., 1998. Extending predictive capabilities to network models. SPE Annual Technical Conference and Exhibition. San Antonio, Texas; 324–336.

Pebesma, E. J., 2004. Multivariable geostatistics in S: the gstat package. *Computers & Geosciences* **30**: 683-691.

Pettersen, O., 2007. Sandstone compaction, grain packing and Critical State Theory. *Petroleum Geoscience* **13**(1): 63–67.

Pringle, J. K., Westerman, A. R., Clark, J. D., Drinkwater, N. J., Gardiner, A.R., 2004. 3D high-resolution digital models of outcrop analogue

study sites to constrain reservoir model uncertainty: an example from Alport Castles, Derbyshire, UK. *Petroleum Geoscience* **10**(4): 343-352.

Pringle, J. K., Howell, J. A., Hodgetts, D., Westerman, A. R., Hodgson, D. M., 2006. Virtual outcrop models of petroleum reservoir analogues: a review of the current state-of-the-art. *First break* **24**: 33-42, March 2006.

Rogiers, B., Mallants, D., Batelaan, O., Gedeon, M., Huysmans, M., Dassargues, A., 2012. Estimation of hydraulic conductivity and its uncertainty from grain-size data using GLUE and artificial neural networks. *Mathematical Geosciences* **44**(6): 739-763.

Rogiers, B., Beerten, K., Smeekens, T., Mallants, D., Gedeon, M., Huysmans, M., Batelaan, O., Dassargues, A., 2013a. The usefulness of outcrop analogue air permeameter measurements for analysing aquifer heterogeneity: Quantifying spatial variability in outcrop hydraulic conductivity. *Hydrological processes*, DOI: 10.1002/hyp.10007.

Rogiers, B., Beerten, K., Smeekens, T., Mallants, D., Gedeon, M., Huysmans, M., Batelaan, O., and Dassargues, A., 2013b. Derivation of flow and transport parameters from outcropping sediments of the Neogene aquifer, Belgium. *Geologica Belgica* **16**(3): 129-148.

van Ranst, E., De Coninck, F., 1983. Evolution of glauconite in imperfectly drained sandy soils of the Belgian Campine. *Zeitschrift für Pflanzenernährung und Bodenkunde* **146**(4): 415–426.

Chapter 5

High resolution hydraulic conductivity logging of borehole cores using air permeability measurements

Based on Rogiers B, Winters P, Huysmans M, Beerten K, Mallants D, Gedeon M, Batelaan O, Dassargues A. High resolution hydraulic conductivity logging of borehole cores using air permeability measurements. Hydrogeology Journal, submitted.

Abstract

Saturated hydraulic conductivity (K_s) is one of the most important parameters determining groundwater flow and contaminant transport in both unsaturated and saturated porous media. This paper investigates the hand-held air permeameter technique for high resolution hydraulic conductivity determination on borehole cores using a spatial resolution of ~0.05 m. We test the suitability of such air permeameter K_s measurements to improve the spatial prediction of classical lab based K_s measurements obtained at a much lower spatial resolution (~2 m). K_s measurements were made on ~350 m of borehole cores originating from the Campine basin, Northern Belgium. The heterogeneity in sediments, ranging from sand to clayey sand with distinct clay lenses, resulted in a K_s range of seven orders of magnitude. Cross-validation demonstrated that using air permeameter K_s data as secondary variable and lab based K_s measurements as primary variable increased performance from R^2 = 0.35 for ordinary kriging (lab K_s only) to R^2 = 0.61 for co-kriging. Due to the large degree of small-scale variability detected by the air permeameter, the spatial granularity in the predicted lab K_s also increases drastically. Air permeameter measurements on borehole cores provide a very cost-effective way to improve spatial predictions of traditional lab based K_s.

5.1. Introduction

Saturated hydraulic conductivity (K_s) is one of the most important parameters determining groundwater flow and contaminant transport in both unsaturated and saturated porous media (*e.g.* Freeze & Cherry 1979). Determining the small-scale variability of this parameter is key to evaluate implications on the use of effective parameters at the scale of groundwater flow modelling applications, typically several orders of magnitude larger than the measurement scale (*e.g.* Ronayne *et al.*, 2010; Huysmans & Dassargues 2009, 2011). Moreover, for stochastic simulations of groundwater flow and even more so for contaminant transport, accurate models on the spatial variability of K_s are very much needed (Nilsson *et al.*, 2007; de Marsily *et al.*, 2005).

Sampling of borehole cores is the most direct way to obtain "hard" small-scale K_s data. It does require expensive minimally disturbed coring, but the obtained data quality is unmatched by other approaches. While several well-established laboratory methods exist for determining K_s, investigating the small-scale variability remains a challenge as it requires collecting large numbers of samples. Indeed, if several hundreds of metres of borehole core have to be hydraulically characterized at the decimetre to centimetre scale, typically several hundreds to thousands of K_s measurements are required, which makes it very costly and time-consuming if traditional methods were used. Hence, the commonly achievable spatial resolution obtained with this approach is limited to the metre scale at best (Rasmussen *et al.*, 1993; Beerten *et al.*, 2010; Yu *et al.*, 2013). To increase spatial granularity in K_s, various geotechnical and wireline geophysical logging tools have been applied for obtaining high

resolution characterization of sediment hydrogeological properties, including traditional resistivity and gamma ray logging (*e.g.* Huysmans & Dassargues 2005; Jiang *et al.*, 2013), and a range of more advanced methods such as IP and NMR techniques (*e.g.* Slater, 2007; Dlubac *et al*,. 2013). The so-called "soft" data is then empirically related to "hard" K_s data through various statistical and geostatistical approaches, or by using theoretical models that relate the measured data quantities such as sediment electrical resistivity or gamma-radiation to K_s. While such soft data may be of a high spatial resolution, its usefulness to estimate K_s is often limited because the measured quantities do not relate directly to the sediment properties governing K_s and hence correlate only moderately with K_s. Moreover, support volumes can be very different between the high resolution soft data and the low resolution hard K_s data (typically from several tens to a few thousand cm^3, e.g. Yu *et al*,. 2013)

Since the 1960s, air permeameter devices have been increasingly used to characterize sediment properties, including K_s through conversion from air permeability (k_a) measurements (*e.g.* Bradley *et al.*, 1972; Welby, 1981). With reliable air permeameters becoming available from the late 80's, a fast and effective semi-direct method exists to determine K_s (*e.g.* Chandler *et al.*, 1989; Davis *et al.*, 1994). As a result, the use of hand-held air permeameter measurements for determining small-scale K_s heterogeneity has been extensively applied, for instance on natural outcrops (Goggin *et al.*, 1988; Jensen *et al.*, 1994; McKinley *et al.*, 2004, 2011; Goss & Zlotnik, 2007; Rogiers *et al.*, 2013a,b; Chapter 2, 3), in quarries (Thomas 1998; Huysmans *et al.*, 2008; Possemiers *et al.*, 2012), or on soils (Kirkham 1947; Loll *et al.*, 1999; Iversen *et al.*, 2003; Beerten *et al.*, 2012).

To our best knowledge, the application of such hand-held air permeameters directly on borehole cores or slabs taken from cores typically used for geological descriptions has not been reported in the literature. This method has the potential to achieve in an efficient manner high resolution K_s data, without the need for additional core sampling from the borehole cores. Moreover, air permeability measurements require times on the order of seconds to minutes, which is only a fraction of the time needed for a constant head test. In principle, the proposed methodology is applicable to any kind of sediment, but excessively long measurement times limit applications for very low K_s values (from several minutes up to half an hour for K_s values between 10^{-7} and 10^{-10} m/s). Also, sediment pore diameter should not have a similar scale as the air permeameter probe opening (9 mm in our case), which limits application in coarse gravels. This paper therefore investigates the usefulness of the hand-held air permeameter technique for high resolution characterization of K_s on borehole cores. This includes i) investigating the hand-held air permeameter technique for high resolution hydraulic conductivity determination on borehole cores using a spatial resolution of ~0.05 m, and ii) testing the suitability of such air permeameter K_s measurements to improve the spatial prediction of classical lab based K_s measurements obtained at a much lower spatial resolution (~2 m).

5.2. Materials & methods

5.2.1. Retrieval of borehole cores

As a case study, we use approximately 350 m of borehole cores (Figure 5.1, Table 5.1) originating from the Mol/Dessel area in the Campine basin, Northern Belgium. The studied sediments are of Miocene to Pleistocene age, with a marine to continental origin, and consist of sand to clayey sand with distinct clay lenses (Figure 5.2), resulting in a K_s range of seven orders of magnitude. During previous studies, two samples were taken from borehole cores every two meters for performing constant head lab permeameter tests (Beerten *et al.*, 2010). This hard data is now used as a reference for and combined with the air permeameter measurements that are performed with a resolution of five centimetres.

The borehole cores studied in this paper originate from two site characterizations carried out in the framework of the ONDRAF/NIRAS low-level radioactive waste disposal research programme (ONDRAF/NIRAS 2010). A 50-m long borehole core (named Dessel-2) was obtained in the first characterization in 2002 (Mallants *et al.*, 2003) aimed at a detailed local study of all relevant lithostratigraphical layers. The cores were taken with a push corer in 96 mm-diameter PVC tubes of 1-1.1 m length, resulting in minimally disturbed core samples, which normally result in a high core recovery (Beerten *et al.*, 2010). Uncertainties about subsurface knowledge were identified through groundwater flow and contaminant transport

modelling studies including sensitivity analyses (Gedeon & Mallants, 2009; Gedeon *et al.*, 2013). To reduce the uncertainties, a second site characterization was performed in 2008 in an area of ~16 km², including six more cored boreholes down to a depth of 40-50 m using the same coring methodology, labelled as Dessel-3, Dessel-4, Geel-1, Kasterlee-1, Retie-1 and Retie-2 (Beerten *et al.*, 2010).

Upon receiving the cores in the laboratory, core handling started with removal of core slabs from the main core (Figure 5.1a) by sawing the cores longitudinally in a one-to-three proportion (Figure 5.1b); the thinnest slab was used for geological characterization and air permeability measurements (carried out in 2012), while the main core section was used for further core sampling (100 cm^3 rings) for lab K_s measurement (Figure 5.1c,d). As this was done on partially saturated cores, the disturbance of the sediment in the slabs is minimal. The unsampled sections were vacuum packed in PE-Al film awaiting future analysis (Figure 5.1e).

Figure 5.1f shows an overview of all core slabs; a typical approximately 1 m-long core section is shown in Figure 5.2 for each lithostratigraphical unit. The Quaternary unit (Figure 5.2a) consists mainly of aeolian sands with some soil development (Beerten *et al.*, 2012). The Mol Upper Sands (Figure 5.2b) are coarse white, very pure quartz sands; the Mol Lower Sands (Figure 5.2c) belong to the same formation, are very well sorted and do not contain the very coarse fraction apparent in the Mol Upper Sands. The Kasterlee Sands (Figure 5.2d) show a higher fines content (both in the clay and silt fractions), and some presence of the typically green-coloured mineral glauconite (up to one weight %). The main aquitard in the study area is the Kasterlee Clay (Figure 5.2e), which is a very heterogeneous alternation of sand layers and clay lenses. The upper part of the Diest Formation, the Diest Clayey Top (Figure 5.2f) also has an increased fines content, and the Diest Sands (Figure 5.2g) are coarse glauconite-bearing sands (up to 30 weight %) with typical bioturbation structures (white spots in Figure 5.2g). Most boreholes contain all lithostratigraphical units. Only in a few cases, the Diest Sands were not penetrated at the bottom of the borehole.

5.2.2. Saturated hydraulic conductivity measurements

Sampling of the main core sections for determination of K_s was performed with 100 cm³ steel ring samples (Figure 5.1c). For characterizing both horizontal (K_h) and vertical (K_v) conductivity at the 100 cm³-scale, a horizontally and vertically oriented sample was taken at each sampling depth at a separation distance of approximately 10 cm (Figure 5.1d). In total 368 samples were retrieved (see Table 5.1) and K_s was measured in the lab using a constant-head permeameter (Klute 1965). Given the quality of sampling, the uncertainty on the resulting values is reduced to a minimum.

Table 5.1: Overview of the cored sections and number of samples of the different boreholes (Beerten *et al.*, 2010). 100 cm³ samples indicated with a star are located more than 20 cm away from the nearest air permeability measurement. K_h: horizontal hydraulic conductivity; K_v: vertical hydraulic conductivity; k_a: air permeability.

Borehole	Drilling	Cored section (m)	# 100 cm³ K_h	# 100 cm³ K_v	# k_a
Dessel-2	2002	50	38(1*)	46	835
Dessel-3	2008	47	21(3*)	23(1*)	779
Dessel-4	2008	40	17(6*)	21	640
Geel-1	2008	45	21(2*)	23(3*)	794
Kasterlee-1	2008	46	17(6*)	23	709
Retie-1	2008	47	20(6*)	23(2*)	770
Retie-2	2008	41	18(5*)	21(1*)	703
Total		316	152(29*)	180(7*)	5230

** samples not within 20 cm distance of a k_a measurement*

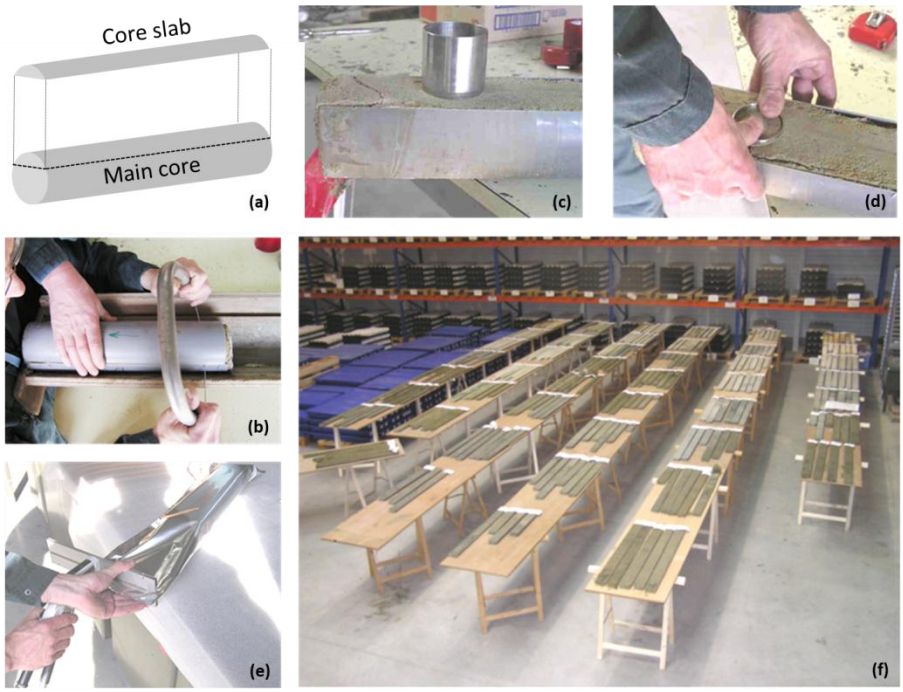

Figure 5.1: Core sampling and preparation for air permeability measurements: a) schematic of slab removal from main core, b) incision is made in main core to the appropriate depth for slab removal, c) 100 cm³ steel core for hydraulic conductivity determination in the lab, d) steel core inserted in core material, e) vacuum sealing of main core for long-term storage, f) display of core slabs for geological description and air permeameter measurements.

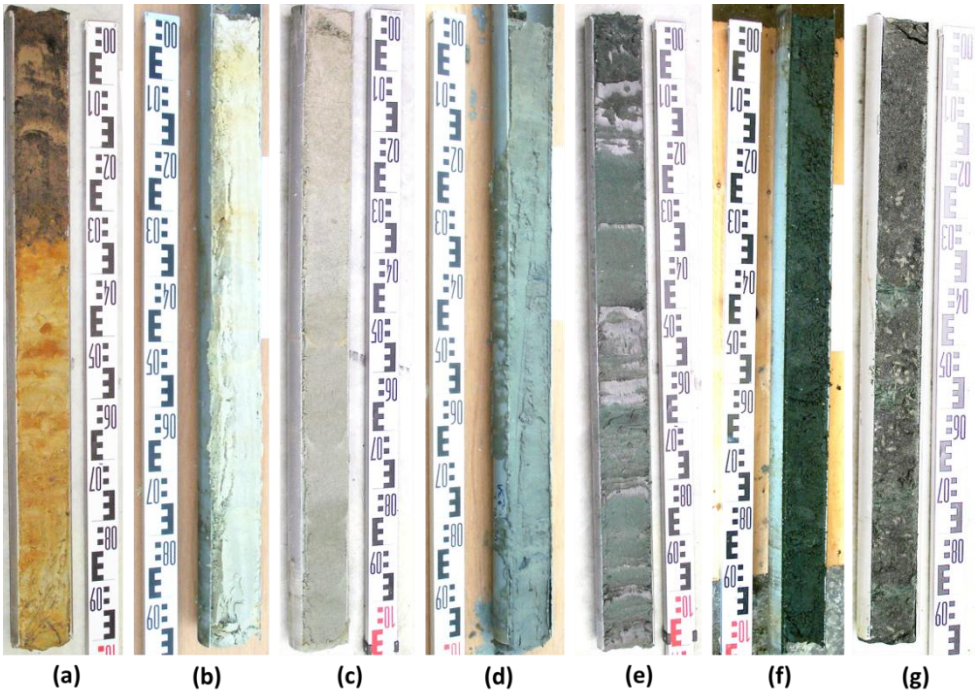

Figure 5.2: Typical borehole cores of the different lithostratigraphical units (from shallowest to deepest sediments): (a) Quaternary; (b) Mol Upper Sands; (c) Mol Lower Sands; (d) Kasterlee Sands; (e) Kasterlee Clay; (f) Diest Clayey Top; and (g) Diest Sands.

The intrinsic sample-scale anisotropy, *i.e.* based on the individual K_h/K_v ratios, was investigated by Beerten *et al.* (2010) using statistical t- and F-tests in comparing both K_h and K_v datasets. Scatterplots of the paired $K_h - K_v$ data are shown in Figure 5.3. The p-values of the t- and F-test were respectively 0.104 and 0.076, for the joint upper aquifer dataset (consisting of the Quaternary, Mol Upper and Lower, and the Kasterlee Sands; Figure 5.3a), indicating that the difference of the mean and variance for the K_h and K_v populations is not significant at the 0.05 (or 5%) level. In other words, the K_h/K_v ratios are not significantly different from one. However, because of the distance between the K_h and K_v samples, the apparent absence of isotropy in K_s might be due to small-scale variability induced by the method of sampling. Figure 5.3a illustrates a rather small coefficient of determination R^2 (0.46). The data points are distributed more or less equally around the 1:1 line ($K_h/K_v = 1$), likely illustrating considerable spatial variability over short distances.

The p-values of the t- and F-test were respectively 0.30 and 0.42, for the Kasterlee Clay aquitard dataset, indicating again that the difference of the mean and variance for the K_h and K_v populations is not significant at the 0.05 (or 5%) level. The scatterplot of the paired $K_h - K_v$ data is shown in Figure 5.3b. The coefficient of determination R^2 is higher than for the upper aquifer data (0.63), but again data points are scattered around the 1:1 line indicating that intrinsic anisotropy might be overprinted by small-scale variability.

For the lower aquifer data (Diest Clayey Top and Diest Sands), the p-values were 0.005 for the t-test and 0.04 for the F-test, indicating that the differences in the mean and variance of the K_h and K_v populations are both significant. In other words, the anisotropy ratio K_h/K_v is significantly different from one. There is however no strong correlation between the K_h and K_v pairs in Figure 5.3c, and many pairs display larger K_v than K_h. Moreover, several outliers indicate that different lithologies were sampled for several pairs of samples. This again indicates the likely importance of the small-scale variability.

As the paired $K_h - K_v$ data anisotropy ratios are not systematically close to one, it cannot be confirmed that the lab based K_s values are isotropic at the sample-scale, despite the fact that the upper aquifer and aquitard marginal distributions show no significant differences in mean and variance between K_h and K_v. Moreover, the marginal distributions of the lower aquifer data differ significantly. Therefore, all further analyses are done on the K_h and K_v datasets separately.

5.2.3. Air permeametry

A Tinyperm II air permeameter device (New England Research & Vindum Engineering, 2011; Figure 5.4a) was used for the borehole cores; several successful studies were previously performed using the same device (Huysmans *et al.*, 2008; Possemiers *et al.*, 2012; Rogiers *et al.*, 2013a,b; Chapter 2, 3). The device consists of a vacuum cylinder, pressure transducer, handle and plunger, and a microprocessor and control unit. The permeameter tip is pressed against the borehole core material, and the plunger is depressed to create a vacuum causing air to flow from the unsaturated porous medium into the device where the gas flow rate and pressure are monitored by the pressure transducer and analysed by the microprocessor unit. Using signal processing algorithms, the unsteady state response function is computed and related to the sample k_a. The exact value of k_a can be determined by an equipment specific calibration curve (New England Research & Vindum Engineering, 2011). To prevent loose sand debris being sucked into the device, a custom made metallic screen was fitted at the outlet. This required recalibration of the TinyPerm II device to correct for modifications to the air flow (Huysmans *et al.*, 2008).

The volume of sediment involved in a permeameter measurement for isotropic porous media is defined by a hemisphere two to four times the internal radius of the tip seal (Goggin *et al.*, 1988a; Jensen *et al.*, 1994). The TinyPerm II has an inner tip diameter of 9 mm, resulting in an investigation depth of 9-18 mm, which is small enough for performing reliable measurements on the core slabs that have a maximum material depth of ~30 mm. Measurements in permeable sands typically take a few seconds, less permeable samples take up to a few minutes and clays might take several dozens of minutes.

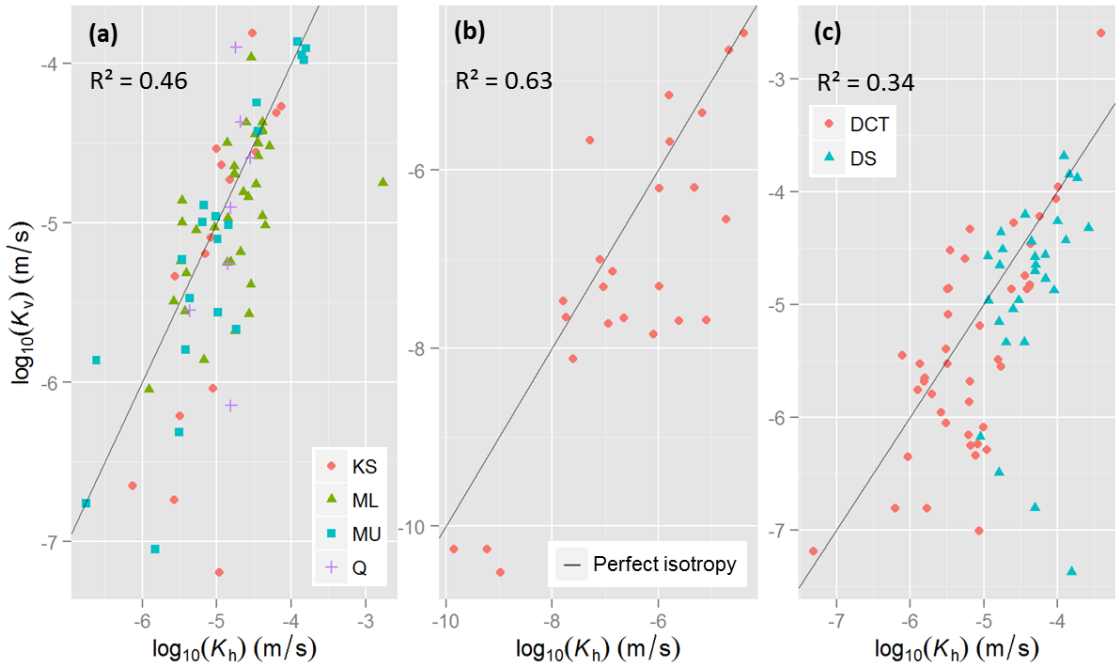

Figure 5.3: Scatterplots of K_h versus K_v derived from laboratory measurements of the 100 cm³ ring samples for: (a) the upper aquifer units Quaternary (Q), Mol Upper (MU) and Lower (ML), and Kasterlee Sands (KS); (b) the Kasterlee Clay aquitard; (c) the lower aquifer units Diest Clayey Top (DCT) and the Diest Sands (DS).

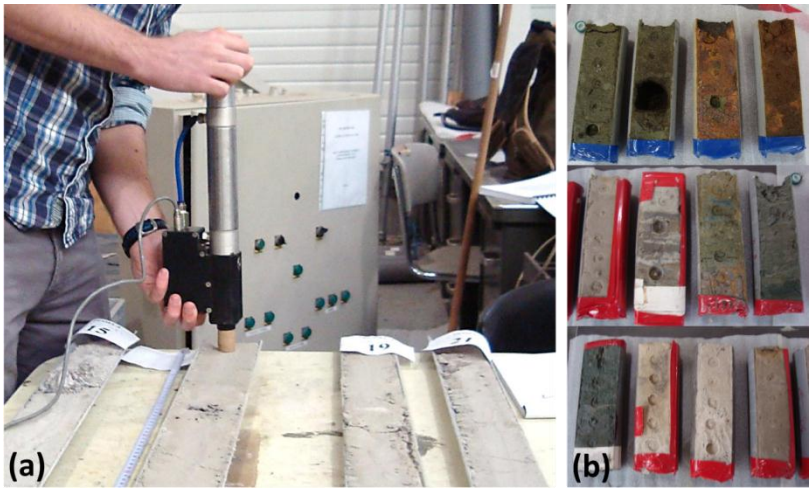

Figure 5.4: (a) Performing air permeameter measurements on borehole core slabs; (b) unpacked core sections for validation of the core-slab-derived values and investigating the impact of moisture content.

Because totally dry sediment conditions are hard to obtain and because of the polar characteristics of water and gas slippage effects, empirical equations have to be used to convert the obtained k_a values into K_s estimates. The empirical equation proposed by Loll *et al.* (1999) is used in this study for the initial k_a-based K_s estimates:

$$K_s = \frac{10^{1.27\times \log_{10}(k_a \times 9.86923e^{-16})+14.11}}{86400} \qquad [5.1]$$

where k_a is expressed in mD, and K_s in m/s. In a recent study on the effect of anisotropy on *in situ* air permeability measurements, Chief *et al.*

(2008) indicated that anisotropy might introduce errors as high as a factor of 2 in air permeability estimates. This is so because air flow lines have different directions in the sediment, and the derived k_a value is somewhere in between the true k_h and k_v parameters. Therefore, we use the air permeameter measurements as secondary data for both the K_h and K_v variables in this study, without making assumptions on the k_a anisotropy or the direction of air flow.

On all 316 approximately 1-m long, air-dried borehole core slabs k_a measurements were carried out with a spacing of approximately 5 cm, resulting in 5230 k_a values. All measurements were performed within 5 days, with one person for handling the air permeameter, and another one to log the Tinyperm II responses.

The measurements are non-destructive, but the structure of very loose sands may be disturbed after the measurement.

5.2.4. Quantification of measurement error and influence of moisture content

Next to the measurements on the core slabs, several additional investigations were performed to better capture the characteristics of the air permeameter device and its limitations. In a first step, the measurement error was quantified by doing 20-30 repeated measurements on each lithology resulting in 280 measurements in total. A second test was designed to investigate the representativity of the core slabs, and the influence of the moisture content on the obtained K_s values. For this, we unpacked 12 of the main cores and separated sections of 30 cm for K_s and gravimetric moisture content measurement. Both properties were measured three times: 1) immediately after unpacking, 2) after drying in air during a week, and 3) after 2 additional days of drying in the oven, resulting in a total of 171 measurements.

5.2.5. Calibration of air permeability-based hydraulic conductivity estimates

All K_s values from the laboratory constant head tests on the 100 cm³ ring samples were paired with the closest k_a-based K_s values (within a maximum distance of 20 cm) to perform a calibration of the detailed high-resolution k_a-based K_s logs. After considering the equation of Loll *et al.* (1999), a site-specific calibration was tested as a means to increase reliability of k_a-based K_s estimates. Moreover, preliminary analysis indicated that the calibration would benefit from including lithostratigraphy as a separate factor in the regression analysis, because different units showed slightly different $k_a - K_s$ relationships. Furthermore, a slight bias was observed between the K_s estimates from the Dessel-2 cores (collected in 2002) and all other cores (collected in 2008); in other words, air permeability measurements were carried out on respectively ten and four year old material. Therefore, we extended the linear model approach and used a linear mixed-effects model with random effects for both the stratigraphy and borehole factors. This was done for K_h and K_v treated separately as predictor variables. The models were fitted with the lme4 package developed in *R* (Bates *et al.*, 2012).

5.2.6. Geostatistical analysis

After calibration of the high-resolution k_a-based K_s measurements, a geostatistical analysis was performed to provide the best possible estimates (*i.e.* spatially interpolated) of the primary K_s data by invoking the spatially cross-correlated secondary k_a data. This analysis consisted of the following steps: 1) experimental variography for the primary and secondary datasets after standardization of the data, 2) fitting of direct variograms and cross-variograms using an intrinsic model of co-regionalization (Goovaerts, 1997), 3) interpolation by co-kriging, and 4) perform a leave-one-out cross-validation to quantify the predictive uncertainty on K_s (kriging variance), as well as the gain in accuracy (performance of the spatial interpolation model) by using the correlated secondary data. These steps were implemented twice, once for the K_h dataset and once for the K_v dataset. The pooled data set (K_h and K_v together) was not considered for this purpose because of its poorer correlation with the K_a-based K_s estimates. All analyses were performed within the R language (R Development Core Team, 2012), making extensive use of the gstat package (Pebesma, 2004).

5.3. Results & discussion

All raw data encompassing the 100 cm³ ring sample data and the k_a-based K_s values obtained from the equation of Loll *et al.* (1999) are shown in Figure 5.5. Especially in the clay-rich units (Kasterlee Clay, Diest Clayey Top) there is

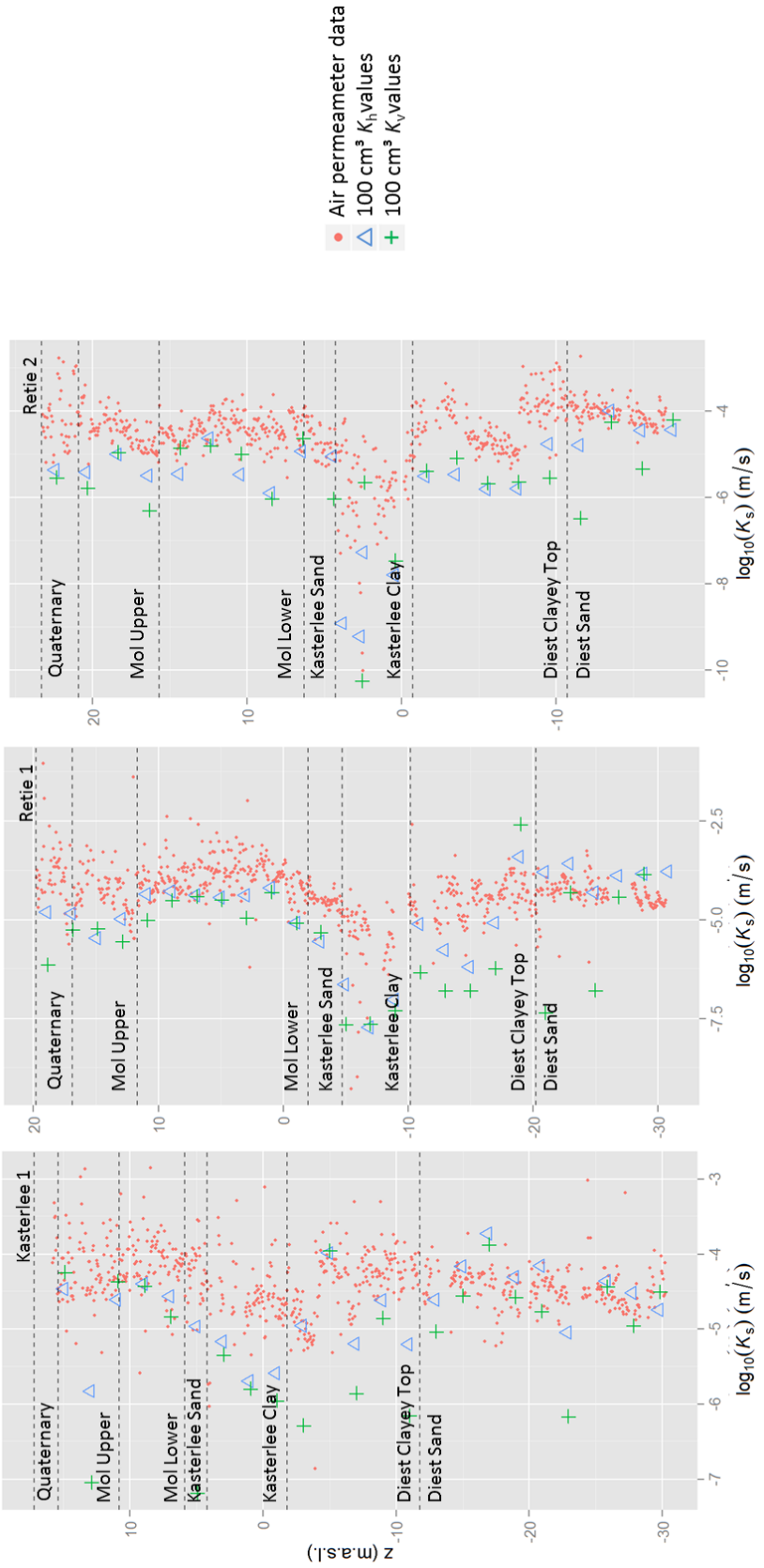

Figure 5.5: Overview of all raw data (100 cm³ lab-based K_s and k_a-based K_s values obtained with the equation of Loll *et al.* (1999)) for the seven boreholes (Dessel-2, 3, 4, Geel-1, Kasterlee-1, Retie-1,2).

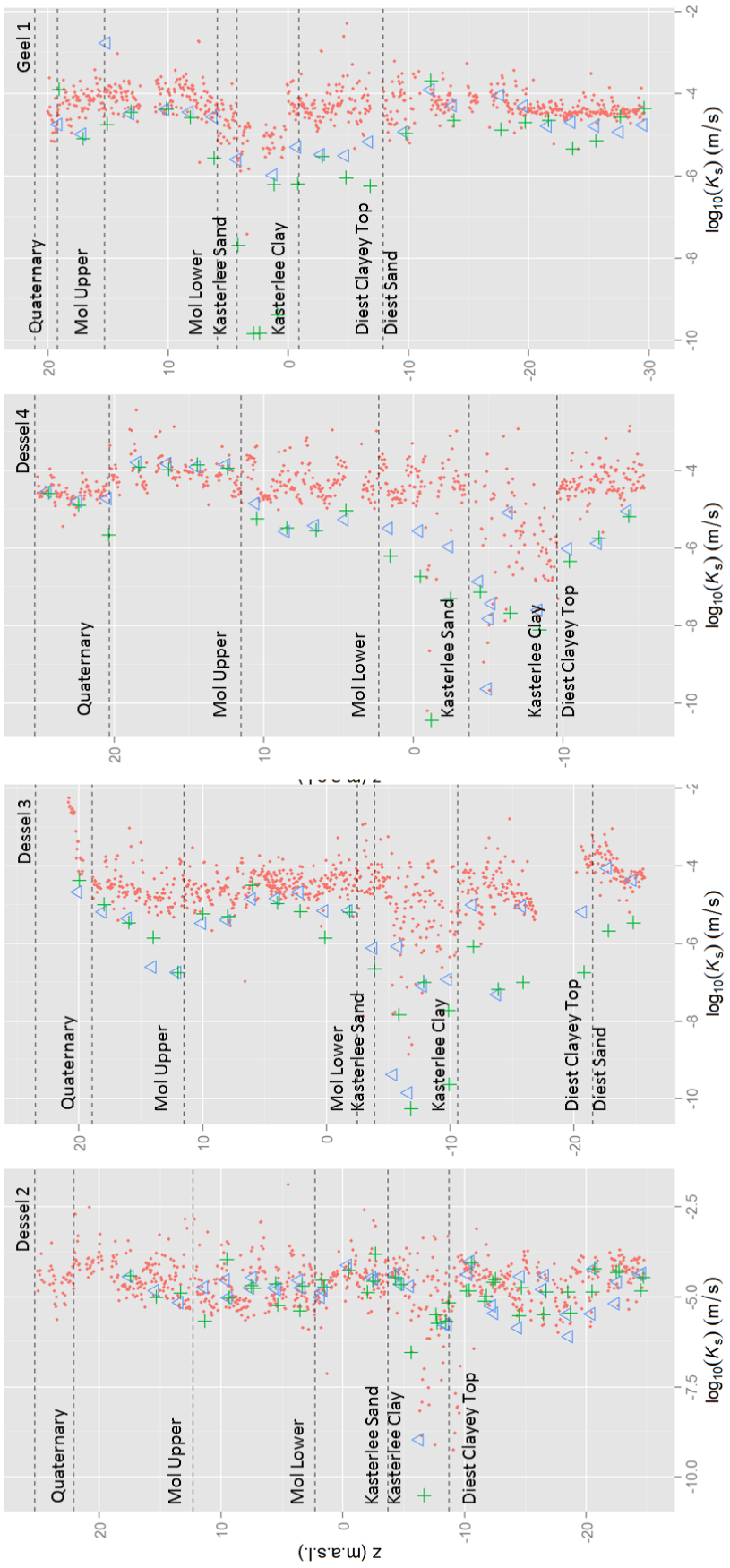

Figure 5.5 (continued).

a clear mismatch between the laboratory-derived and the air permeameter-based values, while in the sand-dominated units a very good match is observed. The sand-dominated units were all very dry during measurement allowing for reliable determination of K_s; presence of clay generally results in less reliable measurements (measurements may not have reached equilibrium, while microscopic fissures in the dried clay may yield overestimations) which may have caused the discrepancies in the clay-rich units. For a given measurement type (K_h, K_v, or the logarithmic k_a-based K_s estimates ($K_{s,ap}$)) the relative differences between the different units seem to be honoured, but the absolute mean K_s values differ significantly from $K_{s,ap}$ for all cases except the Diest Sands K_h. Such discrepancies were not observed when air permeability measurements on outcrop sediments were compared with lab-based K_s values on 100 cm³ ring samples (Rogiers *et al.*, 2013b; Chapter 2). Factors that may have contributed to the discrepancies include borehole core slabs that have been subject to drying in open air and displacement of slabs rendering certain sections in a disturbed condition, thus causing a systematic bias towards higher K_s values. Moreover, the air permeameter data show a considerably higher number of values outside of the 95% confidence interval; this might be due in part to the much larger data set compared to the lab-based data (~14 times), and the smaller support volume of the air permeameter measurements (between 1.5-12 cm^3) compared to the 100 cm³ ring samples.

The measurement error, including the intrinsic variability in response of the device as well as operator-dependent influence (way of handling the plunger), was investigated by doing repeated measurements. The results are shown in Figure 5.7, for three different samples, each time measured by 2 persons. The measurement error clearly is small compared to the variability within a single lithostratigraphical unit (Figure 5.6). The operator influence also shows a relatively small effect, but it is clearly present. It seems to be more important for lower K_s values, but a more systematic analysis might be necessary to confirm this. The extreme high outlier from sample 3 taken by person 1 in Figure 5.7 clearly represents an erroneous measurement, where probably leakage of air between the rubber nozzle of the air permeameter and the sediment occurred. Highly anomalous values like this where discarded when encountered during the borehole core slab measurements, and repeated.

The air permeameter measurements at different times during the drying (*i.e.* decreasing gravimetric water content) of the full-sized cores are shown in Figure 5.8. Water content clearly has an important effect on K_s values, but the differences depend on particle size. For instance, water content changes of a few percent can lead to changes of up to an order of magnitude for the clayey samples. For the sand-dominated units the effect on K_s is small: most values remain within one order of magnitude difference considering a water content range from 16 to zero %. Note however that differences in K_s due to water content changes of a few percent are of comparable magnitude as the measurement error (0.10 for $\log_{10}(K_{s,ap})$).

Table 5.2: Mean and variance for the 100 cm³ log K_h and K_v values and the log k_a-based K_s estimates ($K_{s,ap}$).

Lithostratigraphy	Mean $K_{h,lab}$	Mean $K_{v,lab}$	Mean $K_{s,ap}$	Var $K_{h,lab}$	Var $K_{v,lab}$	Var $K_{s,ap}$
Quaternary	-4.83*	-5.21*	-4.21	0.07*	0.48	0.49
Mol Upper	-4.99*	-5.04*	-4.31	0.67*	0.86*	0.23
Mol Lower	-4.78*	-4.94*	-4.31	0.29	0.21	0.27
Kasterlee Sands	-5.14*	-5.91*	-4.51	0.38	3.01*	0.43
Kasterlee Clay	-6.87*	-7.21*	-5.26	2.53*	3.24*	1.17
Diest Clayey Top	-5.15*	-5.37*	-4.45	0.52	0.89*	0.36
Diest Sands	-4.34	-4.94*	-4.31	0.17	0.74*	0.11

**Significantly different from $K_{s,ap}$*

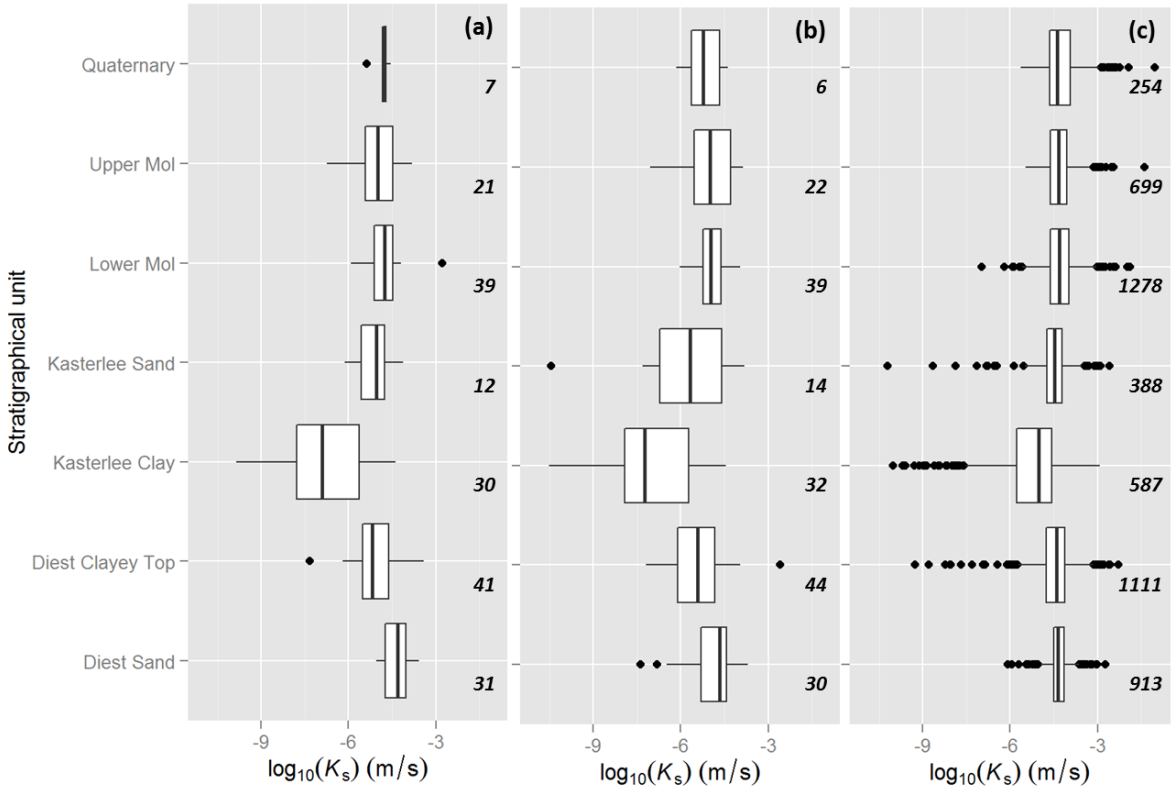

Figure 5.6: Boxplots of median, first and third quartile, and 95% confidence interval K_s values from a) K_h lab analysis, b) K_v lab analyses, and c) air permeameter K_s estimates (based on Loll et al., 1999), for each of the stratigraphical units. The whiskers extend to the most extreme data point which is no more than 1.5 times the interquartile range from the box. The number of observations is provided each time as well.

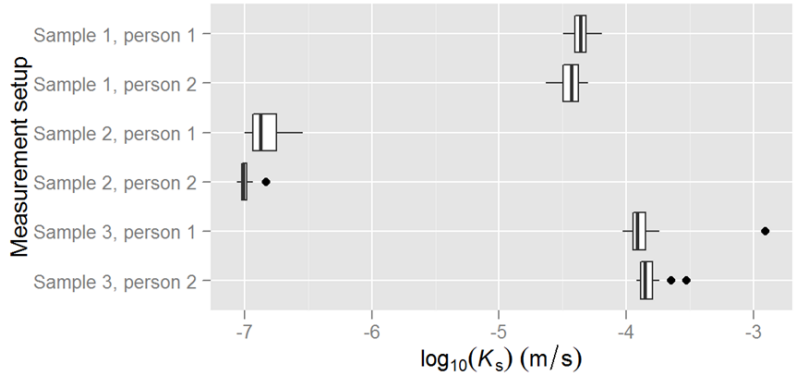

Figure 5.7: Repeated measurements and operator influence. Samples 1 and 3 were measured 30 times, sample 2 20 times due to larger measurement time.

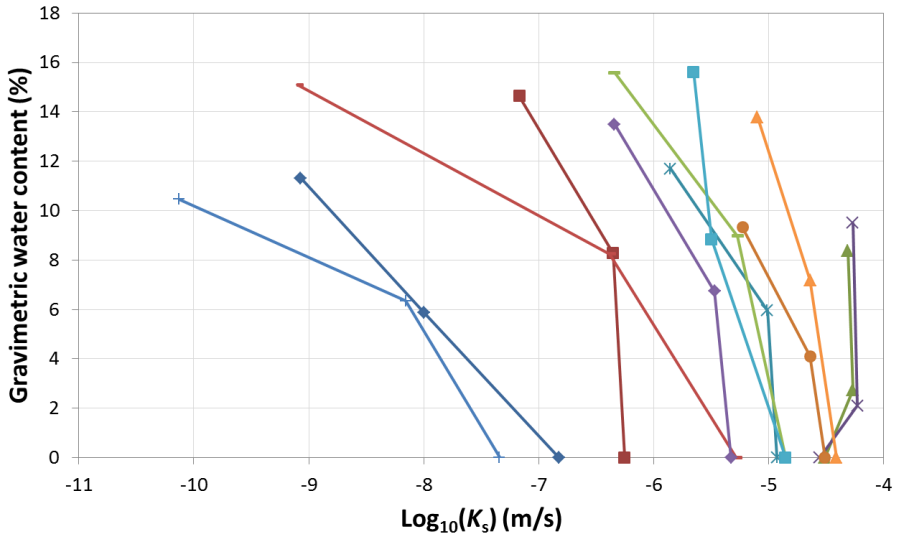

Figure 5.8: Effect of gravimetric water content on air permeability based K_s using full-sized core samples. The highest water content represents the condition after unpacking of the cores (*i.e.* removal of the vacuum seal), the second point is water content after a week of drying in open air, and the final (*i.e.* zero) water content is obtained after 48 hrs drying in an oven.

Because all measurements were performed in the same dry conditions, effects (and bias) within the same unit are similar. Because the k_a-based K_s values are only used as secondary data after calibration with lab-based K_s, effects of water content become unimportant, recognizing though that the degree of correlation between primary and secondary data may depend on water content.

The predictive capacity of the equation of Loll *et al.* (1999) to generate reliable k_a-based K_s estimates is demonstrated in Figure 5.9a. The scatterplot shows that values derived for the predominantly sandy sediments have the correct order of magnitude, while the very low K_s pertaining to Kasterlee Clay clearly shows a systematic bias of two to three orders of magnitude. Overprediction of the air permeameter K_s values for the clayey sediments is likely due to modifications of the pore structure after a long period of drying, also recognizing that the very low K_s values are probably beyond the measurement range of the Tinyperm II, and that an inherent variability is introduced by not having air permeameter and lab measurements at the same locations. The R^2 based on the combined data set (K_h and K_v) is 0.50, but depends on the maximum distance used for matching lab based and air permeameter measurements (*e.g.* using 10 cm rather than 20 cm as maximum distance results in an R^2=0.55). The predictive capacity of the air permeameter K_s estimates can be significantly improved through calibration. Using linear mixed effect models for the K_h and K_v data increases the R^2 to 0.72 (Figure 5.9b).

After standardization of both the lab and calibrated air permeameter data (using the linear mixed effects models for K_h and K_v), experimental variograms were calculated (Figure 5.10). As the data shows considerable scatter within lag distances less than 0.4 m, we decided to extrapolate the semivariance at 0.4 m to a nugget value of zero at the origin, using a spherical model with a 0.4 m range. The increase of the semivariance between 0.4 and 20 m was captured by adding a second spherical model with a range of 12 m. The only differences between the K_h and K_v variogram models are the partial sills, with the K_v model showing the higher semivariance at short distances.

For determining the cross-variograms between lab K_s and air permeameter measurements, we multiplied the partial sills of the direct variogram models with the respective correlation coefficients for K_h (0.88) and for K_v (0.82). This approach is more robust than using the experimental cross-variogram, given the limited amount of primary data (368) in comparison with the secondary data (5230).

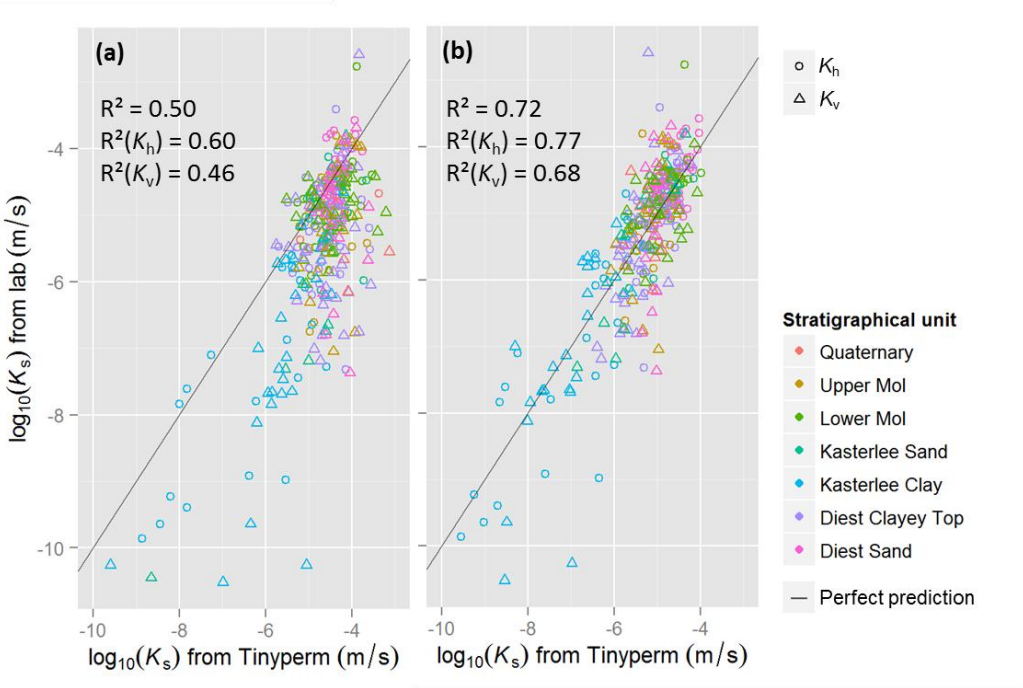

Figure 5.9: a) Scatterplot of the uncalibrated air permeameter K_s estimates using Loll *et al.* (1999) versus laboratory K_s. b) Calibrated k_a-based K_s estimates by means of linear mixed effects models for K_h and K_v, with a random effect for the lithostratigraphy and borehole factors.

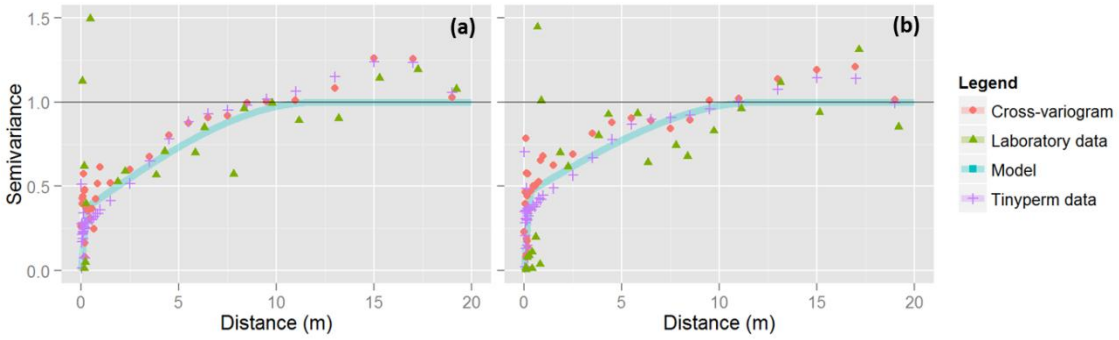

Figure 5.10: Experimental direct variogram and cross-variograms of the standardized data, for a) K_h and b) K_v. The corresponding variogram models consist of a short (0.4 m) and long range (12 m) spherical variogram model.

Table 5.3: Leave-one-out cross-validation results. IDW: inverse distance weighting; OK: ordinary kriging; OCK: ordinary co-kriging; MSE: mean squared error; MAE: mean absolute error; ME: mean error; ρ: correlation coefficient; R²: coefficient of determination; NSeff: Nash-Sutcliffe efficiency.

Performance measure	IDW	OK	OCK
MSE	1.20	1.04	0.68
MAE	0.73	0.68	0.59
ME	0.05	0.02	0.02
ρ	0.56	0.59	0.78
R²	0.31	0.35	0.61
NSeff	-0.06	-0.25	0.57

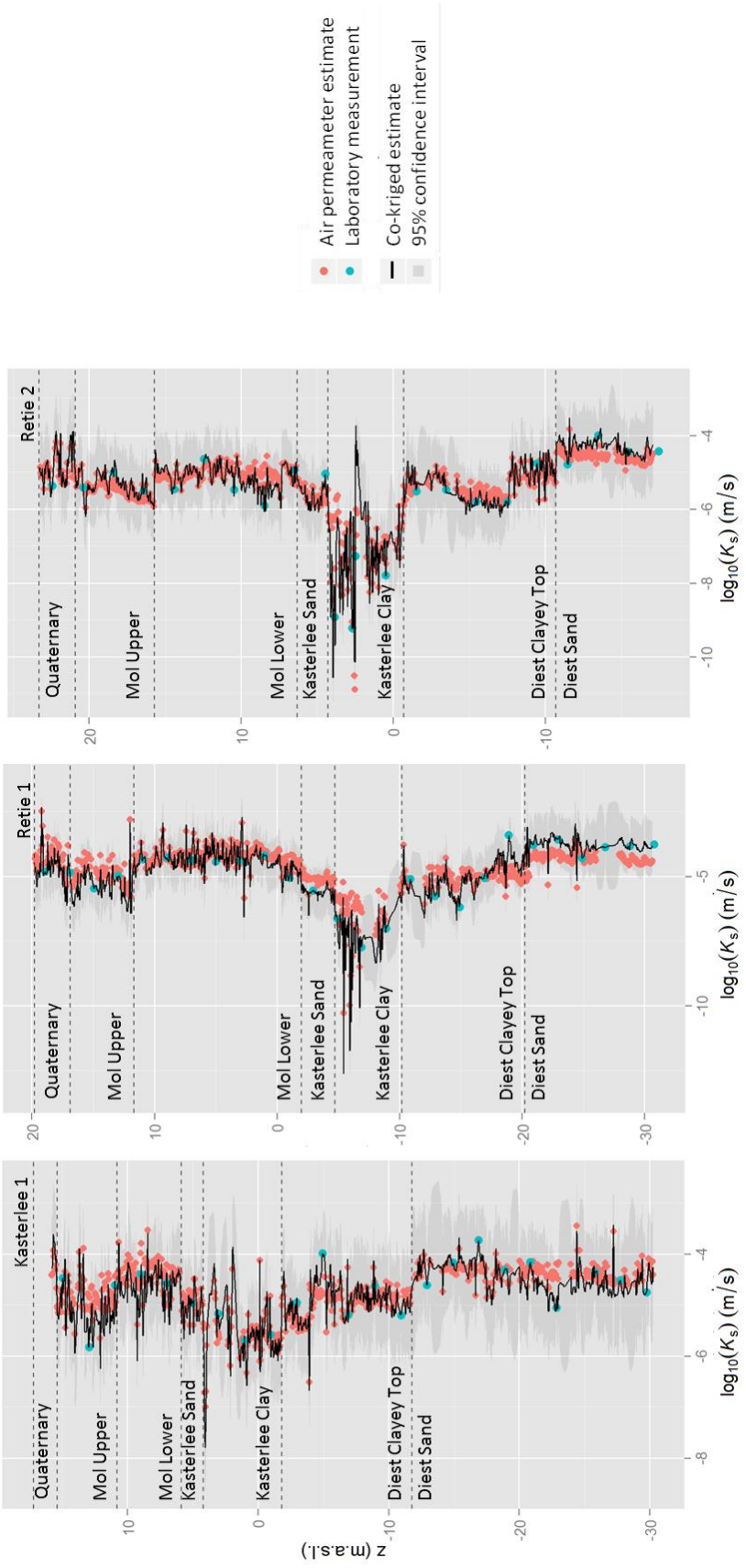

Figure 5.11: Vertical profiles of lab-based K_h values, the calibrated air permeameter-based K_h estimates, and the co-kriging estimates and their 95% confidence interval.

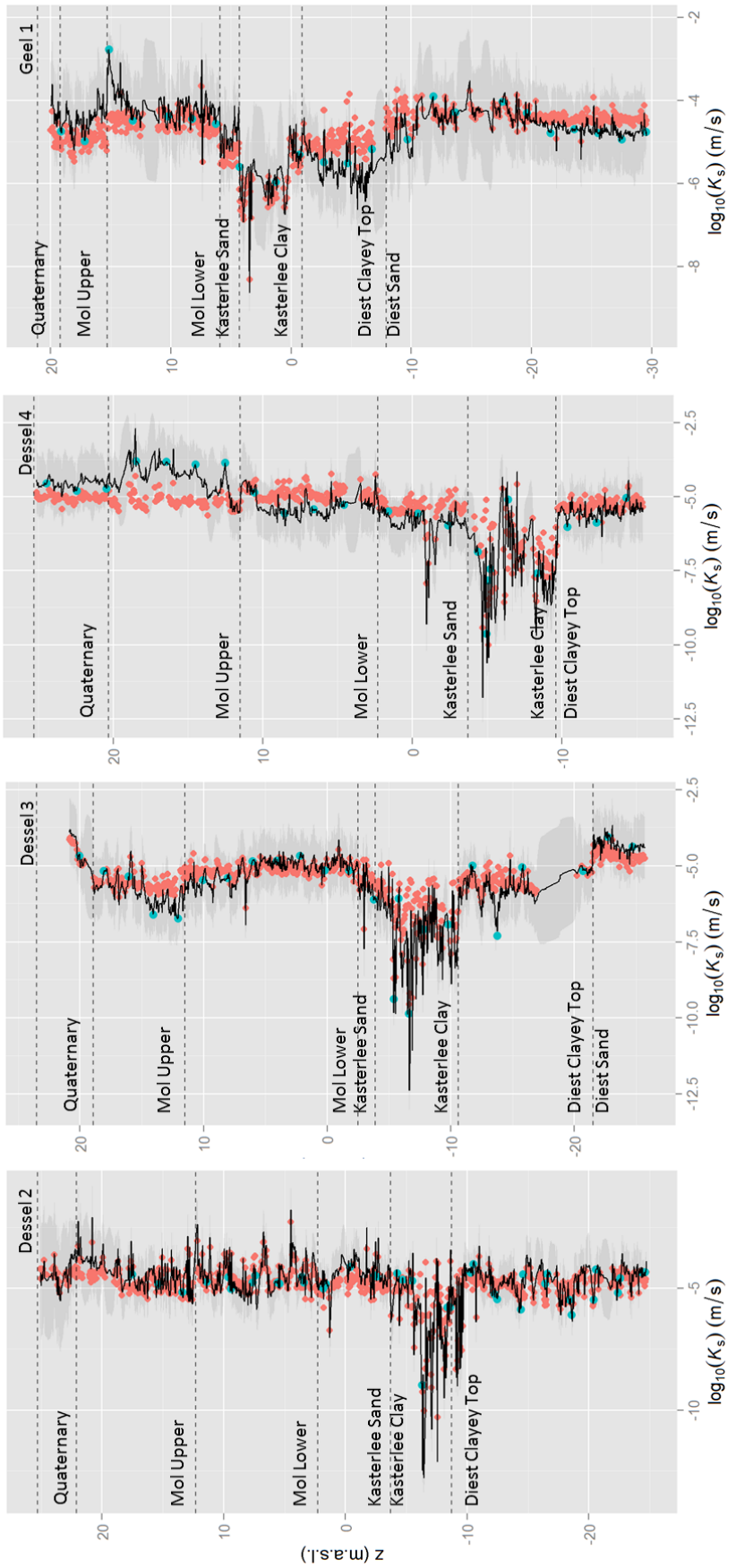

Figure 5.11 (continued).

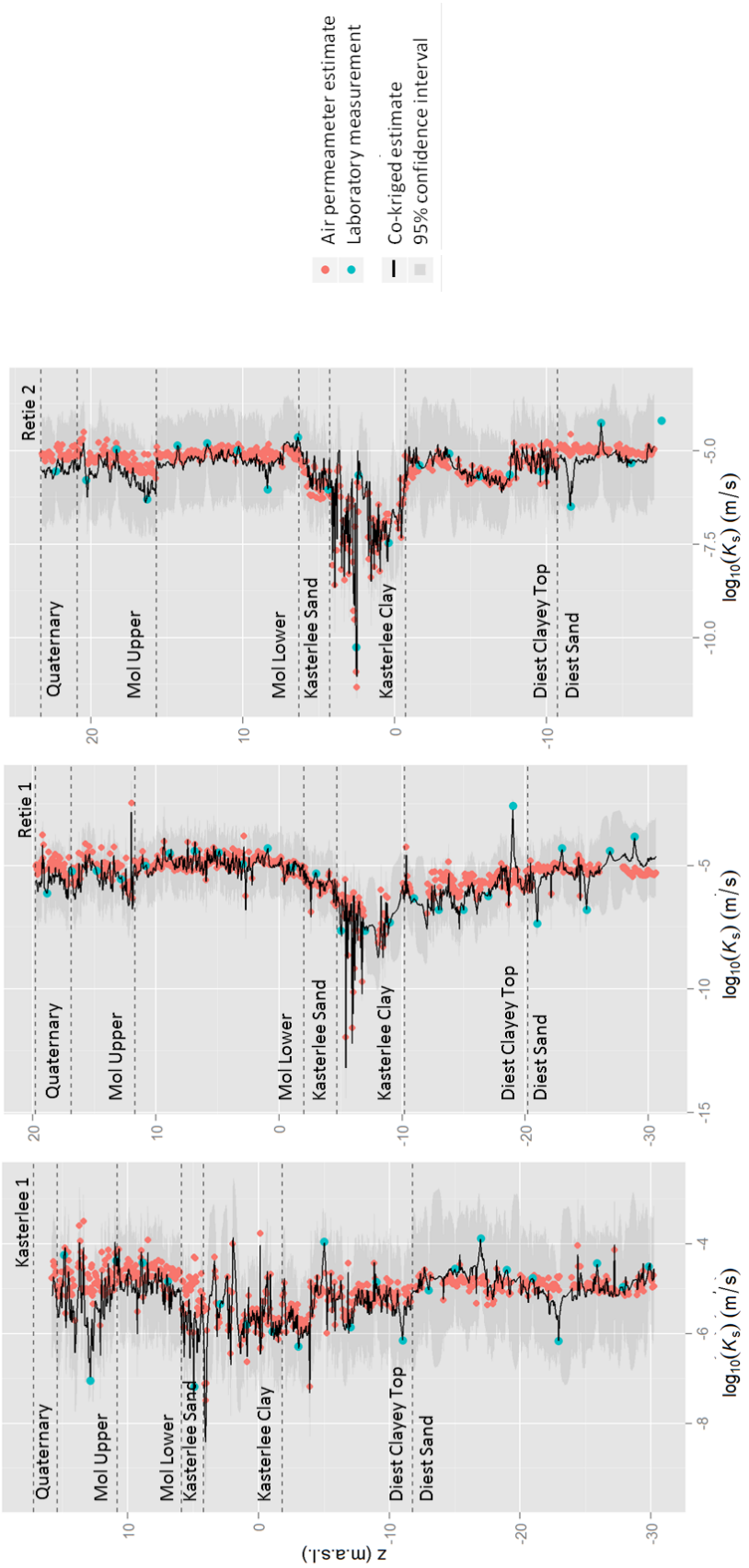

Figure 5.12: Vertical profiles of lab-based K_v values, the calibrated air permeameter K_v estimates, and the co-kriging estimates and their 95% confidence interval.

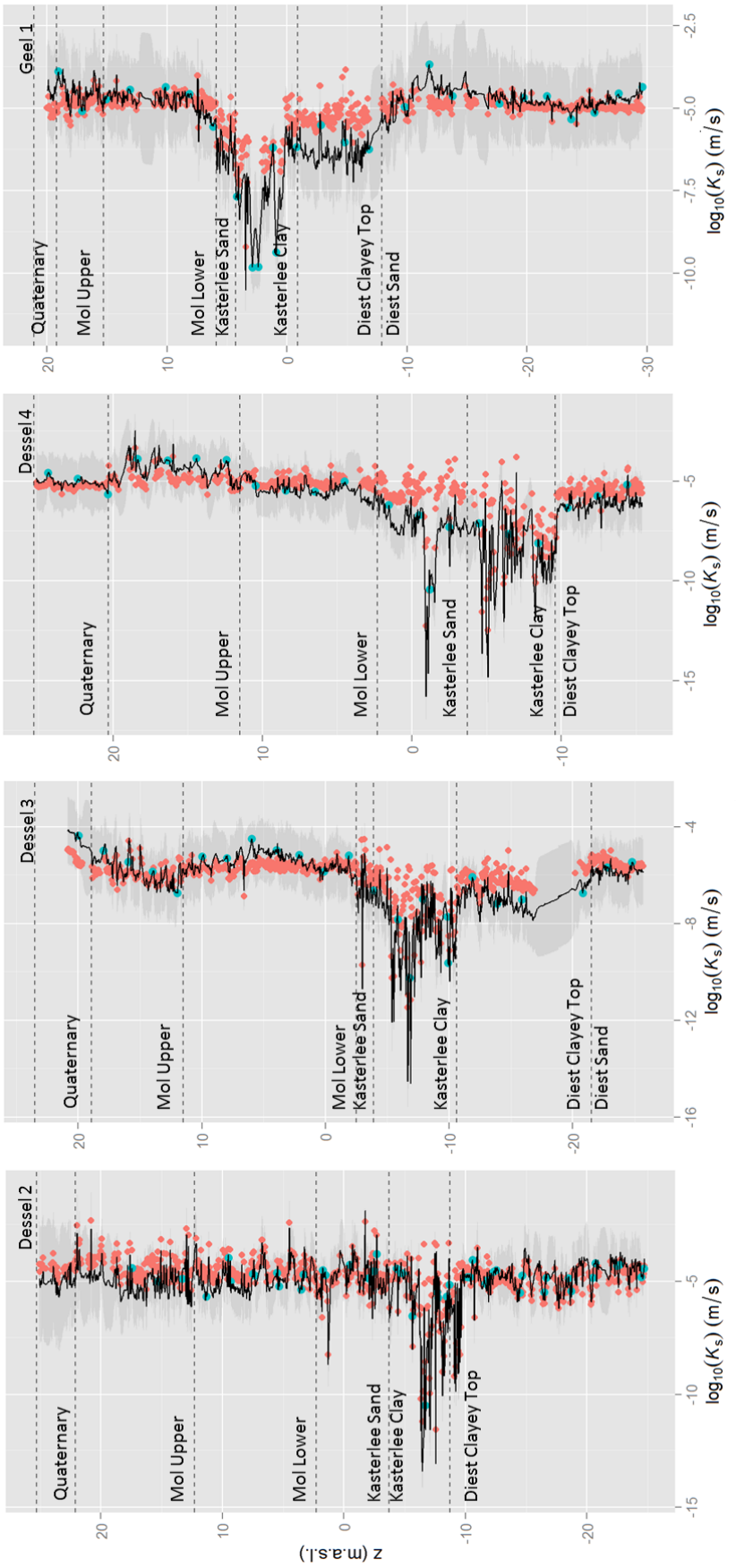

Figure 5.12 (continued).

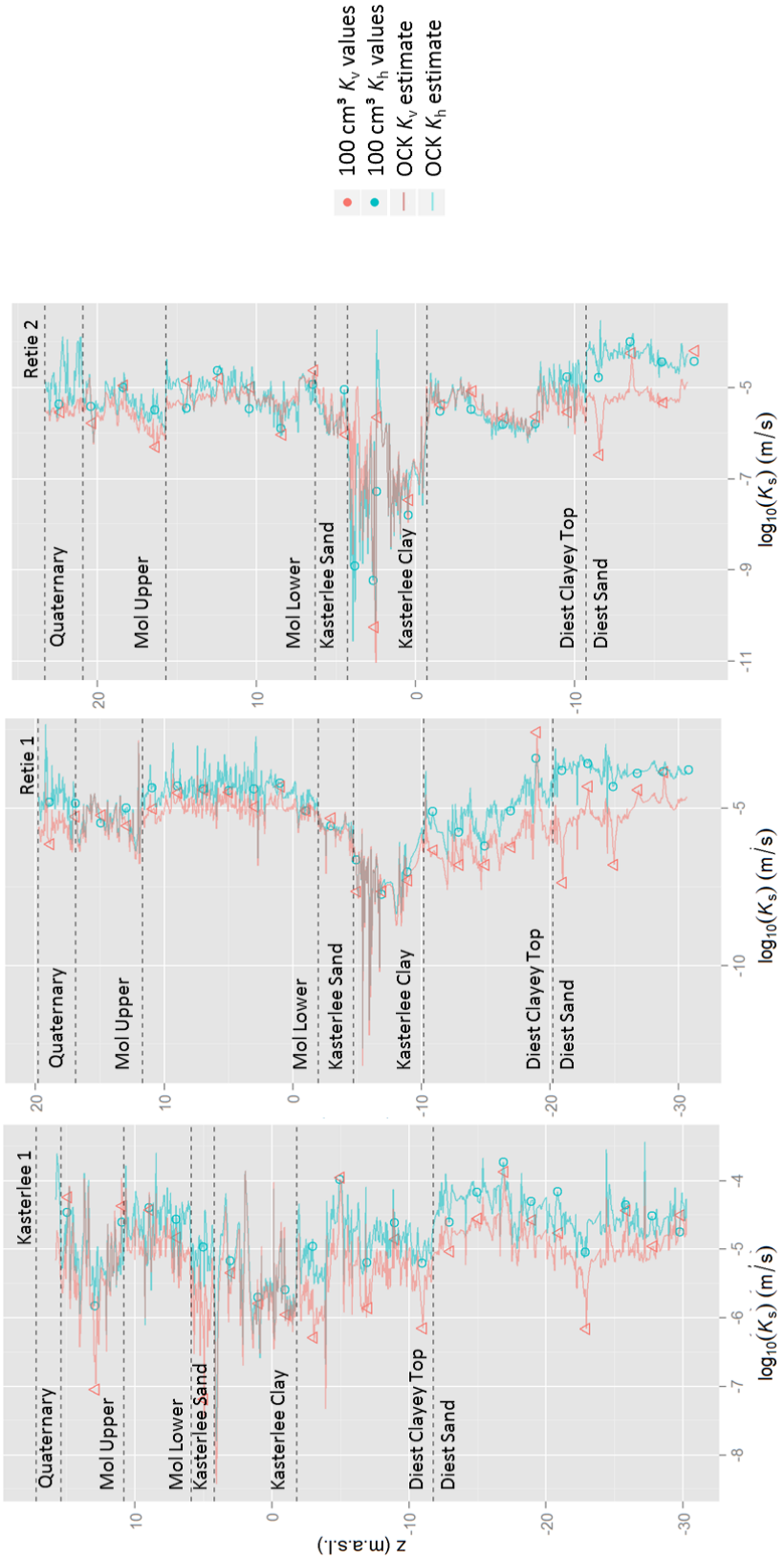

Figure 5.13: Overview of the primary K_h and K_v data together with the corresponding co-kriging estimates.

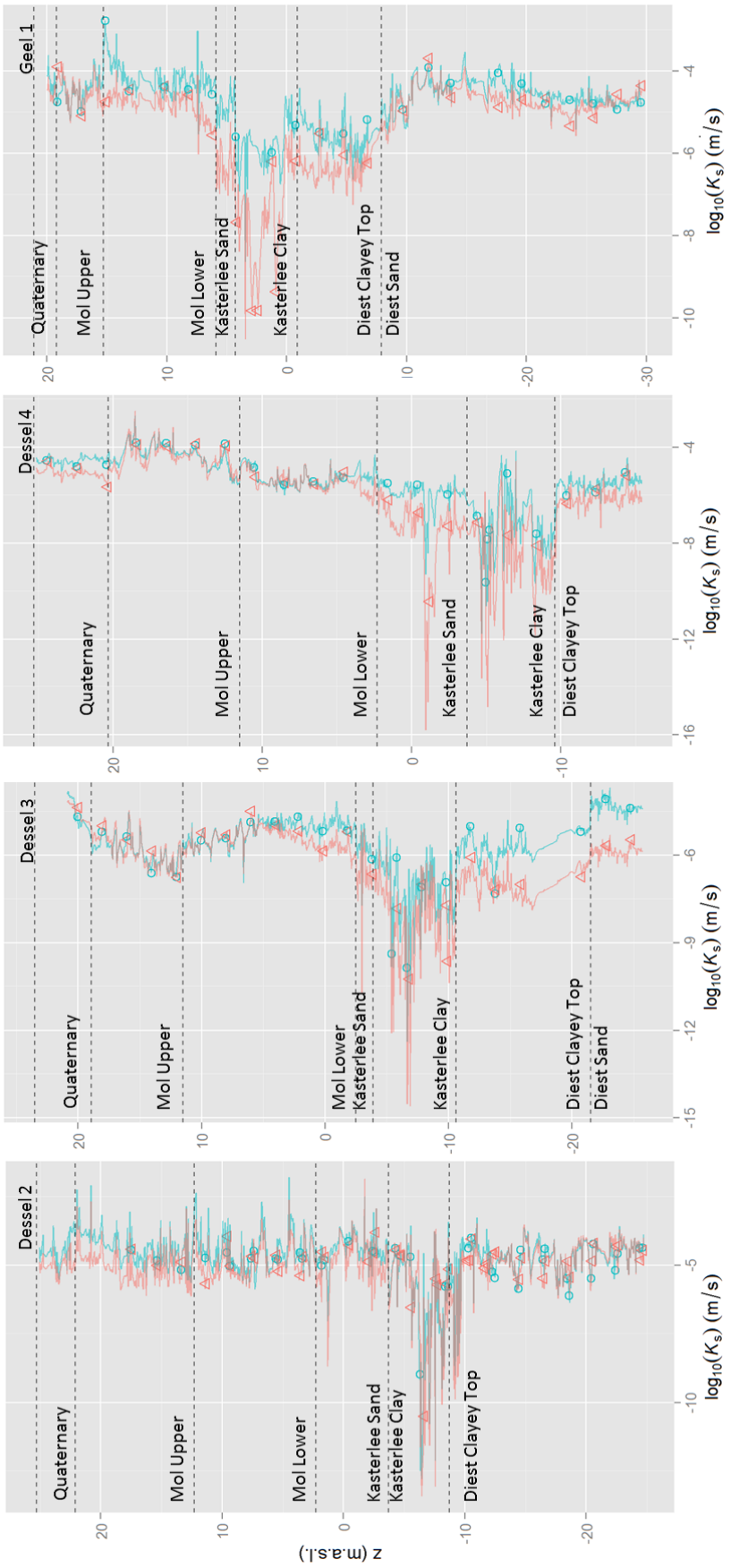

Figure 5.13 (continued).

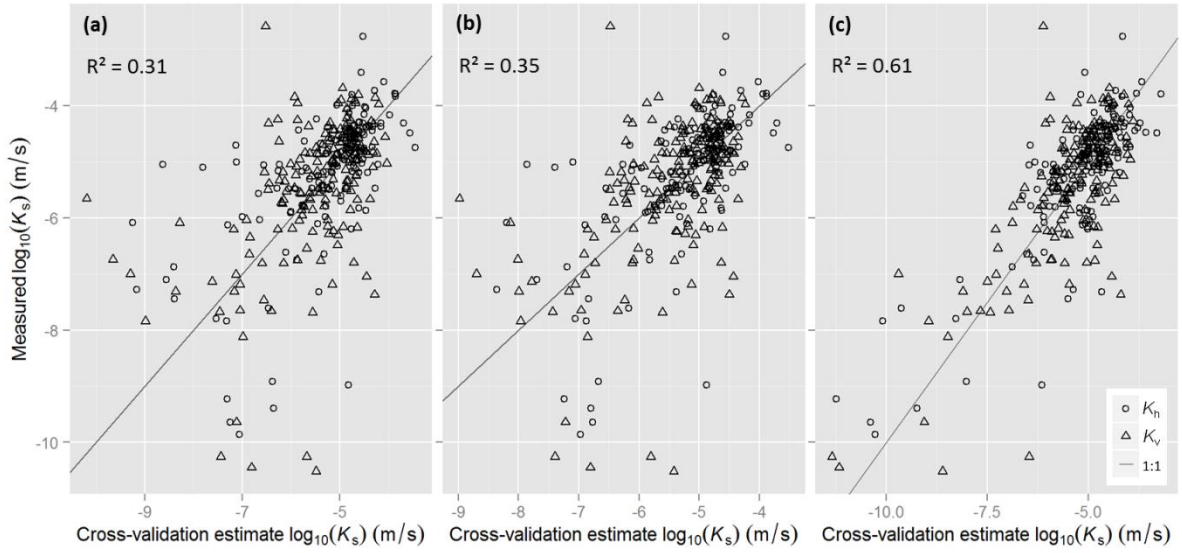

Figure 5.14: Leave-one-out cross-validation results for a) inverse distance weighting, b) ordinary kriging, and c) ordinary co-kriging. R^2 is provided on the figures, other performance measures can be found in Table 5.3.

The best K_h and K_v estimates for each 2 cm along the borehole core sections were then determined using co-kriging with the lab data as primary variable and the air permeameter data as secondary data. The results are displayed in Figure 5.11 and Figure 5.12, together with the 95% confidence intervals (based on the kriging variance). The largest uncertainties are located in the zones were core slabs were missing, and hence no secondary data is available.

Nearly continuous estimates of K_s are obtained in this way, revealing a lot more heterogeneity of the subsurface than the lab-based K_s dataset only. Especially the structure of the Kasterlee Clay displays a high degree of heterogeneity at each borehole location, with several discrete clay lenses sandwiched in-between coarse sand layers. A few other thin lenses with finer material are revealed as well and occur outside of the current boundaries of the Kasterlee Clay unit. Such information could be used to revise the location of several litho-stratigraphical boundaries. Compared to all other boreholes, the Dessel-2 borehole shows more scatter in the data, which could be attributed to the longer exposure of the core slabs to air compared to the 2008 cores.

To investigate the intrinsic anisotropy at the measurement scale of 100 cm³ after the secondary k_a data has been accounted for, the co-kriging estimates for both K_h and K_v are shown in Figure 5.13. The results indicate that mainly the lower aquifer sediments and several parts of the Mol Lower and Kasterlee Sands show systematic intrinsic sample-scale anisotropy, as was previously indicated by the statistical t- and F-tests. The Mol Upper Sands are consistently isotropic, except for the upper part in the Dessel-2 borehole.

Validation of the use of calibrated k_a as secondary variable in a co-kriging approach with lab-based K_s as primary variable and the derived variogram models, was done on the basis of leave-one-out cross-validation (see Table 5.3 and Figure 5.14). Ordinary kriging and ordinary co-kriging was compared for all cases (once for all K_h and once for all K_v samples) to quantify the benefit of using air permeability K_s values as secondary variable for spatial interpolation of K_s. Inverse distance weighting as an alternative interpolation technique was also performed to quantify the benefit of accounting for data-based spatial variability.

According to all performance measures in Table 5.3, the ordinary co-kriging approach with the secondary air permeameter data performs best as spatial interpolation technique for the lab-based K_s measurements. The main gain in performance is for the clayey samples which occur mainly in the more heterogeneous parts and which are more difficult to predict based on

the primary dataset only. Given the large amount of small-scale heterogeneity, the difference in performance between the inverse-distance weighting and ordinary kriging is very small.

5.4. Conclusions

The hand-held air permeameter is a very efficient cost-effective tool to obtain high-resolution information on K_s variability from borehole core slabs. The measurement error was also quantified, and proved to be small compared to the K_s variability. Calibration with independent lab-based K_s measurements improved the predictive capacity of the Loll *et al.* (1999) equation considerably. The regression model for lab-based K_h measurements with linear mixed effects gave the overall best result. Based on a 5-cm measurement interval, the air permeability based K_s values revealed considerable small-scale spatial variability, with an overall range in K_s between 10^{-10} and 10^{-3} m/s.

Spatial interpolation using the site-specific air permeability calibration with linear mixed effects models as secondary variable in an ordinary co-kriging approach proved to be reasonably accurate based on a full leave-one-out cross-validation with an R^2 of 0.61. In comparison, an R^2 of 0.31 and 0.35 was obtained for respectively inverse distance weighting and ordinary kriging. Especially the interpolated K_s estimates of the thin clay lenses improved drastically. Finally, a comparison of the interpolated high-resolution K_h and K_v profiles revealed that at the 100 cm³ sample scale anisotropy is obvious in certain sections, which was not evident from an analysis of the primary lab-based K_s dataset alone.

The analyses were performed for the K_h and K_s datasets separately. Accounting for the correlation between both variables would improve the accuracy of the methodology, but is not straightforward with large small-scale variability and the lack of co-located (at least in vertical direction) samples.

References

Bates, D., Maechler, M., Bolker, B., 2012. lme4: Linear mixed-effects models using S4 classes. R package version 0.999999-0. http://CRAN.R-project.org/package=lme4

Beerten, K., Wemaere, I., Gedeon, M., Labat, S., Rogiers, B., Mallants, D., Salah, S., Leterme, B., 2010. Geological, hydrogeological and hydrological data for the Dessel disposal site. Project near surface disposal of category A waste at Dessel – Version 1.- Brussels, Belgium: NIRAS/ONDRAF, 2010.- 273 p.- NIROND–TR 2009–05 E V1.

Beerten, K., Deforce, K., Mallants, D., 2012. Landscape evolution and changes in soil hydraulic properties at the decadal, centennial and millennial scale: A case study from the Campine area, northern Belgium. *Catena* **95**: 73-84.

Bradley, V. W., Duschatko, R. W., Hinch, H. H., 1972. Pocket permeameter: Handheld device for rapid measurement of permeability: *Bull. American Association Petroleum Geologists* **56**: 568-571.

Chandler, M.A., Goggin, D.J., Lake, L.W., 1989. A mechanical field permeameter for making rapid, non-destructive permeability measurements. *Journal of Sedimentary Petrology* **59**(4): 613-635.

Chief, K., Ferré, T.P.A., Hinnell, A.C., 2008. The Effects of Anisotropy on In Situ Air Permeability Measurements. *Vadose Zone Journal* **7**(3) : 941. doi:10.2136/vzj2007.0164

Davis, J. M., Wilson, J. L., Phillips, F. M., 1994. A Portable Air-Minipermeameter for Rapid In Situ Field Measurements. *Ground Water* **32**(2): 258–266.

De Marsily, G., Delay, F., Goncalves, J., Renard, Ph., Teles, V., Violette, S., 2005. Dealing with spatial heterogeneity. *Hydrogeology Journal* **13**(1): 161–183.

Dlubac, K., R. Knight, Y.-Q. Song, N. Bachman, B. Grau, J., Cannia, J., Williams, 2013. Use of NMR logging to obtain estimates of hydraulic conductivity in the High Plains aquifer, Nebraska, USA, *Water Resour. Res*. **49**: 1871–1886, doi:10.1002/wrcr.20151

Freeze, A.R., Cherry, J.A., 1979. Groundwater. Prentice Hall Inc., New Jersey, 604 p.

Gedeon, M., Mallants, D., 2009. Hydrogeological modelling in support of site characterisation, NIROND-TR 2008-36 E, NIRAS-MP5-03 DATA-LT(HYD), April 2009

Gedeon, M., Mallants, D., Rogiers, B., 2013. Building a staircase of confidence in groundwater modeling: a summary of ten years data collection and model development. Proceedings of MODFLOW and More 2013: Translating Science into Practice. Golden, Colorado, 2-5 June 2013, 6 pp.

Goggin, D.J., Chandler, M.A., Korcurek, G.A., Lake, L.W., 1988. Patterns of permeability in eolian deposits: Page Sandstone (Jurassic), northeastern Arizona. *SPE Formation Eval.* **3**: 297-306.

Goovaerts, P., 1997. Geostatistics for Natural Resources Evaluation (Applied Geostatistics Series) (p. 496). Oxford University Press, USA.

Goss, D., Zlotnik, V.A., 2007. Air permeameter investigation of surficial dune structures in the Nebraska Sand Hills. *AAPG Bulletin* **91**(5): 645–652. doi:10.1306/11200606009

Huysmans, M., Dassargues, A., 2005, Stochastic analysis of the effect of heterogeneity and fractures on radionuclide transport in a low permeability clay layer, *Enviromental Geology* **48**(7): 920-930.

Huysmans, M., 2006. A geostatistical methodology for modeling groundwater flow and transport in low-permeability media: application on Boom Clay, Ieper Clay and Toarcian argillites.

Huysmans, M., Peeters, L., Moermans, G., Dassargues, A., 2008. Relating small-scale sedimentary structures and permeability in a cross-bedded aquifer. *Journal of Hydrology* **361**(1-2): 41-51.

Huysmans, M., Dassargues, A., 2009. Application of multiple-point geostatistics on modelling groundwater flow and transport in a cross-bedded aquifer (Belgium). *Hydrogeology Journal* **17**(8): 1901–1911.

Huysmans, M., Dassargues, A., 2011. Direct Multiple-Point Geostatistical Simulation of Edge Properties for Modeling Thin Irregularly Shaped Surfaces. *Mathematical Geosciences* **43**(5): 521–536. doi:10.1007/s11004-011-9336-7

Iversen, B.V., Moldrup, P., Schjonning, P., Jacobsen, O.H., 2003. Field Application of a Portable Air Permeameter to Characterize Spatial Variability in Air and Water Permeability. *Vadose Zone Journal* **2**(4): 618-626.

Jensen, J.L., Glasbey, C.A., Corbett, P.W.M., 1994. On the interaction of geology, measurement, and statistical-analysis of small-scale permeability measurements. *Terra Nova* **6**(4): 397–403.

Jiang, Z., Schrank, C., Mariethoz, G., Cox, M., 2013. Permeability estimation conditioned to geophysical downhole log data in sandstones of the northern Galilee Basin, Queensland: Methods and application. *Journal of Applied Geophysics* **93**: 43-51. doi:10.1016/j.jappgeo.2013.03.008

Kirkham, D., 1947. Field method for determination of air permeability of soil in its undisturbed state. *Soil Sci. Soc. Am. Proc.* **11**:93–99.

Loll, P., Moldrup, P., Schjønning, P., Riley, H., 1999. Predicting saturated hydraulic conductivity from air permeability: Application in stochastic water infiltration modeling. *Water Resources Research* **35**(8): 2387–2400.

Mallants, D., Labat, S., Gedeon, M., 2003. Bijkomende sitekarakterisatie voor de nucleaire zone Mol-Dessel: bepaling van de hydrogeologische parameters. SCK•CEN-R-3703 (p. 82). Mol, Belgium.

McKinley, J.M., Lloyd, C.D., Ruffell, A.H., 2004. Use of variography in permeability characterisation of visually homogeneous sandstone reservoirs with examples from outcrop studies. *Mathematical Geology* **36**: 761–779.

McKinley, J.M., Atkinson, P.M., Lloyd, C.D., Ruffell, A.H., Worden, R.H., 2011. How Porosity and Permeability Vary Spatially With Grain Size, Sorting, Cement Volume, and Mineral Dissolution In Fluvial Triassic Sandstones: The Value of Geostatistics and Local Regression. *Journal of Sedimentary Research* **81**: 844–858.

New England Research, Vindum Engineering, 2011. TinyPerm II Portable Air Permeameter, User's Manual. Retrieved from http://www.vindum.com/TinyPermManual.pdf on 14-06-2011.

Nilsson, B., Højberg, A., Refsgaard, J., Troldborg, L. 2007. Uncertainty in geological and hydrogeological data. *Hydrology and Earth System Sciences* **11**: 1551-1561.

ONDRAF/NIRAS, 2010. Het cAt-project in Dessel. Een langetermijnoplossing voor het Belgische categorie A-afval. Retrieved from http://www.niras-cat.be/downloads/cAt_masterplan_NL_LOW.pdf on 07-12-2011.

Pebesma, E.J., 2004. Multivariable geostatistics in S: the gstat package. *Computers & Geosciences* **30**: 683-691.

Possemiers, M., Huysmans, M., Peeters, L., Batelaan, O., Dassargues, A., 2012. Relationship between sedimentary features and permeability at different scales in the Brussels Sands. *Geologica Belgica* **15**(3): 156-164.

Rasmussen, T.C., Evans, D.D., Sheets, P.J., Blanford, E.H., 1993 Permeability of Apache Leap Tuff: Borehole and core measurements using water and air. *Water Resources Research* **29**(7): 1997–2006.

R Development Core Team, 2012. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/.

Rogiers, B., Beerten, K., Smeekens, T., Mallants, D., Gedeon, M., Huysmans, M., Batelaan, O., Dassargues, A., 2013a. Derivation of flow and transport parameters from outcropping sediments of the Neogene aquifer, Belgium. *Geologica Belgica* **16**(3): 129-147.

Rogiers, B., Beerten, K., Smeekens, T., Mallants, D., Gedeon, M., Huysmans, M., Batelaan, O., Dassargues, A., 2013b. The usefulness of outcrop analogue air permeameter measurements for analysing aquifer heterogeneity: Quantifying outcrop hydraulic conductivity and its spatial variability. *Hydrological processes*. DOI: 10.1002/hyp.10007

Ronayne, M. J., Gorelick, S.M., Zheng, C., 2010. Geological modeling of submeter scale heterogeneity and its influence on tracer transport

in a fluvial aquifer. *Water Resources Research* **46**(10): 1-9.

Slater, L., 2007. Near Surface Electrical Characterization of Hydraulic Conductivity: From Petrophysical Properties to Aquifer Geometries—A Review. *Surveys in Geophysics* **28**: 169–197. doi:10.1007/s10712-007-9022-y

Thomas, C.J.S., 1998. Reservoir characterization of a shallow marine sandstone: the Lower Cretaceous Sandringham Sands (Leziate Beds) and Carstone formations, eastern England. *Petroleum Geoscience* **4**(3): 215-219.

Welby, C.W., 1981. A technique for evaluating the hydraulic conductivity of Saprolite. UNC-WRRI-81-164

Yu, L., Rogiers, B., Gedeon, M., Marivoet, J., De Craen, M., Mallants, D., 2013. A critical review of laboratory and in-situ hydraulic conductivity measurements for the Boom Clay in Belgium. *Applied Clay Science* **75-76**: 1-12.

Chapter 6

Estimation of hydraulic conductivity and its uncertainty from grain-size data using GLUE and artificial neural networks.

Based on Rogiers B, Mallants D, Batelaan O, Gedeon M, Huysmans M, Dassargues A. 2012. Estimation of hydraulic conductivity and its uncertainty from grain-size data using GLUE and artificial neural networks. Mathematical Geosciences 44(6): 739–763.

Abstract

Various approaches exist to relate saturated hydraulic conductivity (K_s) to grain-size data. Most methods use a single grain-size parameter and hence omit the information encompassed by the entire grain-size distribution. This study compares two data-driven modelling methods, *i.e.* multiple linear regression and artificial neural networks, that use the entire grain-size distribution data as input for K_s prediction. Besides the predictive capacity of the methods, the uncertainty associated with the model predictions is also evaluated, since such information is important for stochastic groundwater flow and contaminant transport modelling.

Artificial neural networks (ANNs) are combined with a generalized likelihood uncertainty estimation (GLUE) approach to predict K_s from grain-size data. The resulting GLUE-ANN hydraulic conductivity predictions and associated uncertainty estimates are compared with those obtained from the multiple linear regression models by a leave-one-out cross-validation. The GLUE-ANN ensemble prediction proved to be slightly better than multiple linear regression. The prediction uncertainty, however, was reduced by half an order of magnitude on average, and decreased at most by an order of magnitude. This demonstrates that the proposed method outperforms classical data-driven modelling techniques. Moreover, a comparison with methods from literature demonstrates the importance of site specific calibration.

The dataset used for this purpose originates mainly from unconsolidated sandy sediments of the Neogene aquifer, northern Belgium. The proposed predictive models are developed for 173 grain-size - K_s pairs. Finally, an application with the optimized models is presented for a borehole lacking K_s data.

6.1. Introduction

The quantification of saturated hydraulic conductivity (K_s; see list of symbols) is important for developing groundwater flow and contaminant transport models to support decision making about environmental impact assessment, groundwater contamination cleanup, water resources evaluation, or site monitoring. As an alternative for measuring K_s, grain-size data has been used often to indirectly obtain K_s, as it is one of the cheapest methods to obtain predictor variables for K_s. Hence, the development of (semi-)empirical methods to provide reliable K_s predictions has received lots of attention since the end of the 19th century. The large amount of literature devoted to developing pedotransfer functions in soil physics sciences indicates that measuring K_s directly is not always the most effective way to obtain the needed information, but there are additional arguments to demonstrate the need for predicting K_s instead of measuring it.

A first argument is that physically disturbed samples of subsurface sediments are much more common then undisturbed samples. Only good quality borehole cores allow for K_s measurements on minimally disturbed sediment.

List of symbols.

Symbol	Explanation
p	Layer number
i;j	Neuron number
$w_{i,j}^{p}$	Weight for output of i^{th} neuron in layer p-1 to input for j^{th} neuron in layer p
b_i^p	Bias i^{th} neuron in layer p
x_i^p	Output i^{th} neuron in layer p
y	Model predictions
Y	Observations
η	Learning rate
δ	Delta error
α	Momentum parameter
μ; d	Learning rate adaptation factors
K_s	Saturated hydraulic conductivity
C_h; $C_{K\text{-}C}$	Hazen and Kozeny-Carman empirical coefficient
L_2; L_{11}; …; L_{1000}	Grain size fractions < 2;11;…;1000 μm
f_i	i^{th} grain size fraction
D_{li}; D_{si}	Largest and smallest grain size diameters of fraction f_i
e	Void ratio
SF	Grain shape factor
s; S	sample number; total number of samples
$L(\theta_i\|Y)$	Likelihood for the i^{th} parameter set conditioned on the observations Y
$\sigma_{\varepsilon,i}^2$	Error variance for the ith model
σ'^2_{ε}	Maximum likelihood estimate of the error variance
σ_{obs}^2	Observed variance
N	Weighting power
n	Number of samples used in model derivation
n_p	Amount of neurons in layer p

Grain-size analysis can however still be performed on disturbed core samples. A pedotransfer function developed in the previous setting gives the possibility to make detailed K_s estimates, with a corresponding uncertainty, when undisturbed samples are absent. Since disturbed sediment is easier and less expensive to retrieve than undisturbed borehole cores, this makes the characterization of small-scale hydraulic conductivity much more efficient.

A second argument is that grain size data is definitely more ubiquitous than K_s data. If one thing is measured in the research sector or the industry concerning unconsolidated sediments, it is the grain size distribution. This is the most basic characteristic of the sediment, and is therefore very important to know when correlating different lithological units in space. The ability to make reliable predictions of K_s, starting from grain size, therefore allows for gathering much more information on the hydrogeology. This is especially true if former studies, or studies with different goals where performed at a certain site or for a certain sediment.

One of the earliest predictive K_s models is the Hazen Equation (Hazen 1892), which expresses the hydraulic conductivity (or permeability) as function of the square of the 10^{th} percentile grain-size diameter (d_{10}):

$$K_s = C_h d_{10}^2 \qquad [6.1]$$

with the empirical coefficient C_h. This coefficient however ranges over three orders of magnitude in literature (Carrier 2003). Numerous attempts have been made since Hazen to relate hydraulic conductivity and

grain-size distribution for soil and sediments. In their review, Cronican and Gribb (2004) mention the use of following grain-size parameters: d_{10}, geometric mean grain-size diameter (d_m), sand, silt and clay percentages, d_{50}-d_{10}, and the x-intercept of the d_{50}~d_{10} straight line. Vienken and Dietrich (2011) recently evaluated several existing methods in which d_{20}, d_{50}, d_e (effective diameter) and the coefficient of uniformity $U = d_{60}/d_{10}$ are used. An attempt to correlate K_s to grain-size data for hydrogeological applications for clay-rich sediments is reported by Wemaere *et al.* (2008).
One of the main limitations of these (semi-)empirical formulations is that they all use one or very few parameters from the available grain-size distribution. While certain cumulative percentages or particle sizes might yield a reasonable correlation for a given hydraulic conductivity dataset, the remaining information encompassed by the entire grain-size distribution is often omitted. A typical example of a predictive model that includes information from the entire grain-size distribution through the use of an effective particle diameter (d_e), is the Kozeny-Carman Equation (Kozeny 1927; Carman 1938, 1956):

$$K_s = C_{K-C}\left(\sum f_i / \left(D_{li}^{0.404} \times D_{si}^{0.595}\right)\right)^{-2} \times (1/SF^2)\,(e^3/(1+e)) \qquad [6.2]$$

with the empircal coefficient $C_{\text{K-C}}$, grain size fractions f_i, with largest and smallest grain size diameters D_{li} and D_{si}, a grain shape factor SF, and the void ratio e. Because there are still limitations associated with the latter model (Carrier 2003), approaches that allow "calibration" of the K_s - grain-size distribution relationship have been developed. Van De Genachte *et al.* (1996) used principal component analysis in combination with multiple linear regression to model infiltration parameters, including K_s, of sandy soil. Another example of such development is the use of artificial neural networks (ANNs) for modelling soil hydraulic properties including K_s as function of grain-size distribution and possibly other material parameters such as bulk density, organic matter and carbonate content (*e.g.* Schaap *et al.*, 1998; Pachepsky *et al.*, 1999).
ANNs are computational tools that have found extensive application in a wide range of research areas. ANNs can be seen as a form of nonlinear regression (Basheer and Hajmeer 2000). Moreover, it has been proven theoretically that multilayer feedforward networks are universal approximators (Hornik *et al.*, 1989; Hornik 1991). Environmental applications of this technique include streamflow, flood and rainfall forecasting (*e.g.* Kişi 2007; Jain and Kumar 2007; Tiwari and Chatterjee 2010; Valverde Ramírez *et al.*, 2005), processing remote sensing data (Linderman 2004), flow and transport simulations (*e.g.* Morshed and Kaluarachchi 1998), groundwater table prediction (*e.g.* Coppola *et al.*, 2005; Joorabchi *et al.*, 2009) and water quality modelling (*e.g.* Zou *et al.*, 2009; Khalil *et al.*, 2011). In soil science, ANNs have been used extensively for predicting soil hydraulic properties from more easily measurable parameters. Schaap and Leij (1998) and Schaap *et al.* (2001) used ANNs for predicting soil water retention and log-transformed soil hydraulic conductivity from limited to extended sets of basic soil properties. The possible input variables encompass sand, silt and clay percentages, bulk density and retention points at 10 or 33 kPa. A hierarchical set of practical models available in the ROSETTA software was created for different levels of data availability (Schaap *et al.*, 2001). Minasny *et al.*, (2004) used sand, silt, clay content, bulk density, and saturated water content in an ANN prediction of soil hydraulic functions measured by multistep outflow data. Nakhaei (2005) used 8 cumulative particle size fractions for predicting log-transformed hydraulic conductivity for loamy sand soils with ANNs, with individual modelling of different soil types being superior to joint modelling. Agyara *et al.*, (2007) used besides grain size and bulk density also organic carbon and 9 different terrain parameters to predict K_s.
As discussed above, most efforts of applying ANNs for predicting K_s are situated within the soil sciences, and use mainly clay, silt and sand content as grain-size input variables. There are fewer applications of the method that use data from sedimentary aquifers, possibly due to data scarcity which results from limited accessibility. Moreover, in the framework of stochastic modelling and risk assessment, the quantification of uncertainty related to these predictions is of equal importance, and this aspect seems to be addressed rarely. Uncertainty estimation is however not standard practice in ANN modelling. Asefa (2009) and Tiwari and Chatterjee (2010) show that using random input and network parameter selection in combination

with traditional ANN modelling allows to obtain uncertainty estimates. The framework that is developed here to produce in combination reliable predictions and their uncertainties, makes use of the generalized likelihood uncertainty estimation (GLUE) approach (Beven and Freer, 2001). The principle of GLUE is based on a rejection of the concept of the optimal model. Instead, multiple models are considered to be equally acceptable if their performance, which is assessed through a likelihood measure, is alike. A quantitative measure of model performance is thus used to assess the acceptability of each model and to weigh each model accordingly in the model ensemble prediction and uncertainty estimates.

The objective of this paper is to develop an ANN ensemble modelling approach to predict hydraulic conductivity with multiple grain-size fractions as data input, and to provide reasonable uncertainty estimates using the GLUE methodology. The resulting GLUE-ANN hydraulic conductivity predictions and associated uncertainty estimates are compared with the ones obtained from linear regression models by using a leave-one-out cross-validation approach. An example application is shown in which hydraulic conductivity predictions are made for locations where only grain-size data was available. The used K_s and grain-size data originate from the Neogene aquifer in northern Belgium, which consists of several mostly unconsolidated sublayers of quartz-rich sands with hydraulic conductivity values ranging between 10^{-3} and 10^{-11} m/s, the latter from clay lenses within an aquitard.

6.2. Methods & materials

6.2.1 Measurements

6.2.1.1 Core sampling

The study area covers approximately 60 km^2 in the north of Belgium and is part of the Nete basin (Figure 6.1). At the study site the top 11-26 m of the aquifer are mainly composed of i) Quaternary deposits (2 – 5 m thick) with grain-size varying from fine to coarse sand, ii) well-sorted fine, white quartz sands usually referred to as Mol Sands that become more greenish medium sand in the western part of the study area, and iii) underlying Kasterlee Sands that contain up to one weight % glauconite (Beerten *et al.*, 2010). A second, much thicker aquifer exists under a very heterogeneous aquitard, the latter between 5 and 7 m thick and referred to as Kasterlee Clay. Stratigraphically the top of the second aquiffer mainly consists of very coarse Diest Sands (~84 m thick in the Dessel 5 borehole), with glauconite up to 30 weight %. The Dessel, Berchem and Voort (DBV) sands are present below the Diest sands, overlying the gradual transition to the Boom Clay, which forms a regional aquitard (Wemaere *et al.*, 2008). The total thickness of the DBV layers amounts to around 70 meters at the study site.

Nine boreholes were drilled to depths ranging from 50 (6 boreholes), 90 (2 boreholes) to 200 m. Contiguous coring was undertaken for all 50 m and 90 m long cores and for the lower section of the 200 m long borehole (from 90 to 200 m). A pushing corer was used, developed for unconsolidated sediments, resulting in high recovery and low disturbances. The undisturbed cores were retrieved in 1.1 m long and 96 mm inner diameter PVC tubes and stored for further analyses (Beerten *et al.*, 2010).

The reconnaissance boreholes with a depth of 50 m (Dessel 3, Dessel 4, Dessel 5, Geel 1, Kasterlee 1, Retie 1 and Retie 2) and 200 m (Dessel 5) were all used for model development, with hydraulic conductivity measurements being available for each of them. For the two 90 m deep boreholes (named GT 01 and GT 02), only grain-size data are available. They will be used in a final application of K_s prediction using the proposed GLUE-ANN model. Both reconnaissance and 90 m long boreholes penetrate the upper aquifer, the Kasterlee Clay aquitard, and the lower aquifer.

Sampling consisted of sawing the 1.1-m long PVC cores longitudinally in two unequal parts (approximately in a 20/80 % division of the core diameter). The upper longitudinal section was used for geological description, while the lower part was used for sampling before being vacuum packed for storage awaiting future analysis. Cylindrical plugs (100 cm^3) were pressed in the cores for hydraulic conductivity (horizontal and vertical K_s) measurements and grain-size analysis (Beerten *et al.*, 2010).

6.2.1.2 Grain-size measurements

Grain-size measurements were performed by means of two different techniques: a detailed analysis using a sedigraph yielding 20 size fractions (~0.5, 0.7, 1, 1.4, 2, 3, 4, 6, 8, 11, 16, 22, 32, 44, 63, 88, 125, 177, 250 and > 250 µm), and a standard method (European standard EN

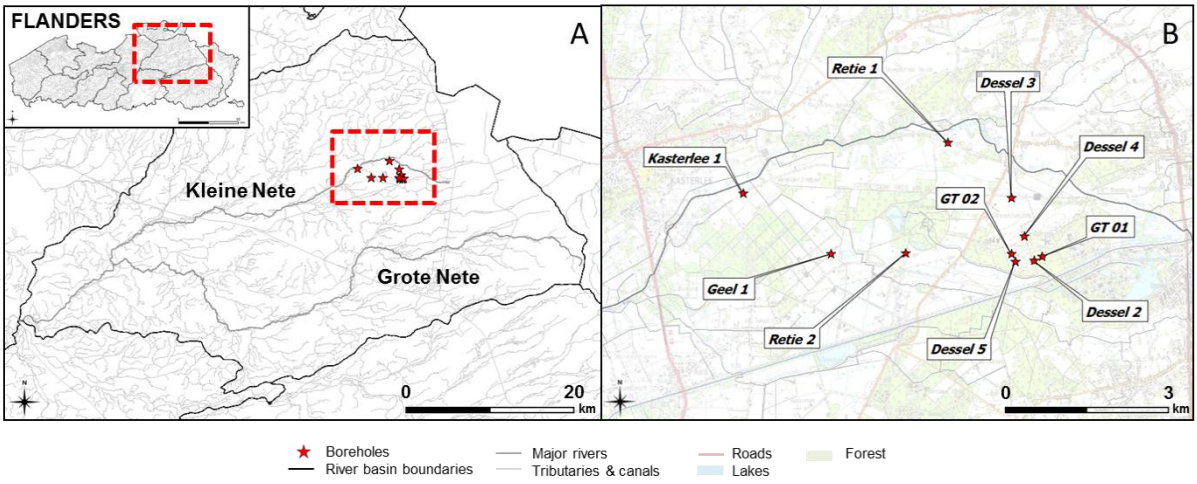

Figure 6.1. A) Location of the study area in the Nete basin and Flanders, and B) the 10 borehole locations.

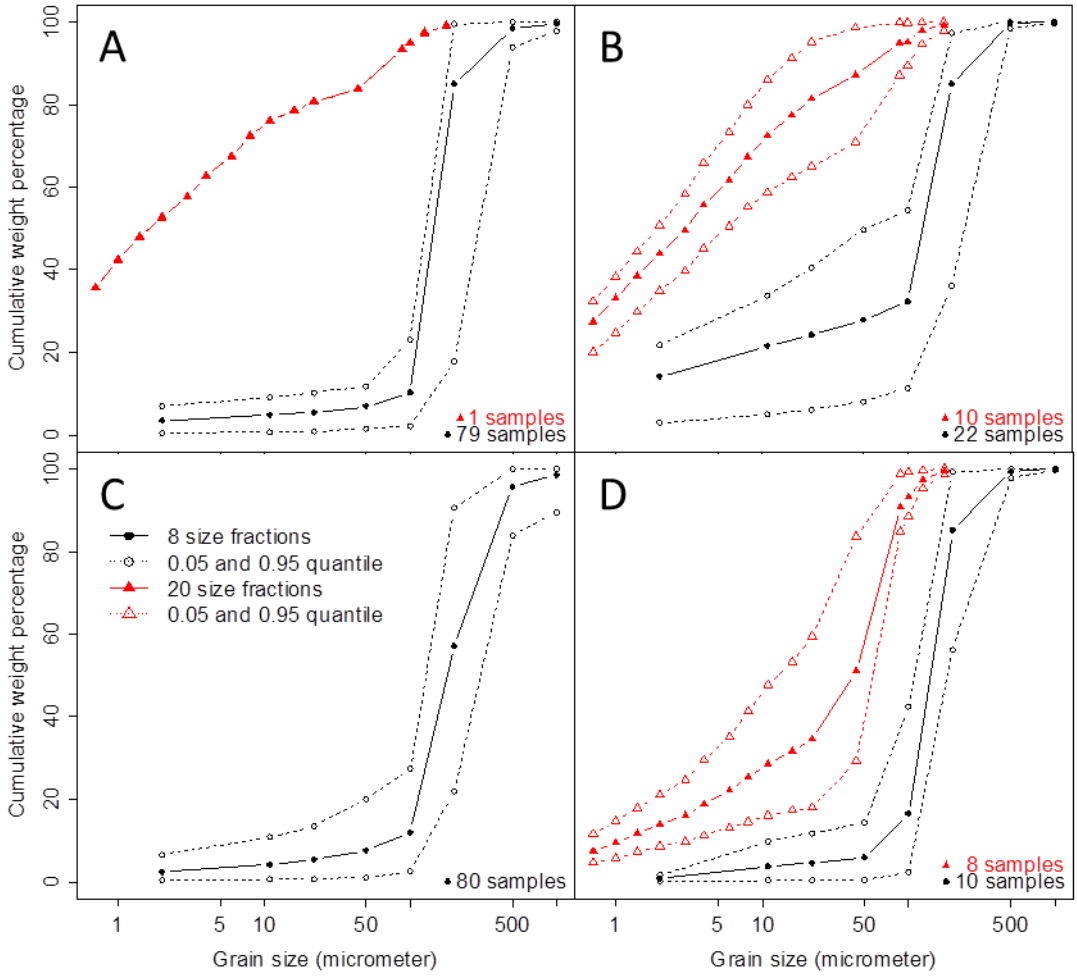

Figure 6.2. Grain-size distributions of the upper aquifer (A), the Kasterlee Clay aquitard (B), the lower aquifer (C), and the deepest part (90-200 m) of the lower aquifer (D). For C only sand samples (8 size fractions) were analysed. In the other graphs, the clay samples (20 size fractions) were plotted separately.

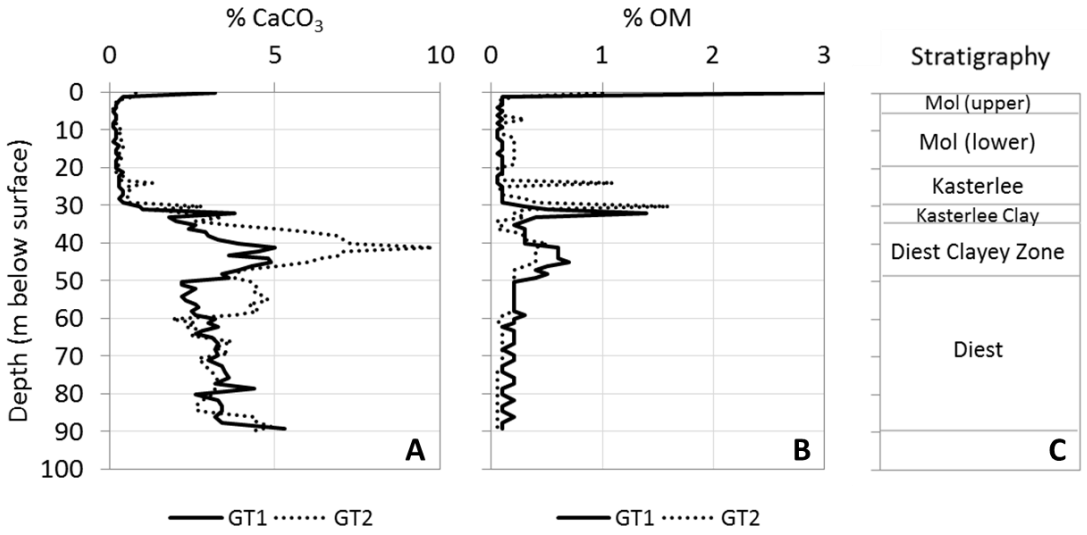

Figure 6.3. A) Carbonate content, B) organic matter, and C) stratigraphy for the two geotechnical boreholes (GT1 & GT2).

933-1) yielding 8 size fractions (2, 10, 20, 50, 100, 200, 500 and 1000 µm). The former technique was applied to clayey samples, while the latter was applied to the sandy samples (see Figure 6.2).

Grain-size analyses with the sedigraph method were performed after eliminating carbonates by acidification, organic matter by oxygenation and particles larger than 250 µm by sieving. Characterization of the particles is based on the sedimentation while the quantification is based on the attenuation of an X-ray beam. Because of the higher resolution in the fine particle ranges (from ~22 to 0.5 µm) it was specifically applied to the clayey samples (19 in total).

Grain-size analyses of the sand samples have been performed according to the European standard EN 933-1. Samples were prepared by eliminating carbonates and organic matter as was done for the sedigraph method. For the > 50 µm particles standard sieving was used with mesh sizes of 100, 250, 500, and 1000 µm. For the < 50 µm particles, a peptising agent was added first after which the solid/water mix was put in a suspension cylinder for determination of the fine particles with class boundaries of 2, 11, and 22 µm.

All particle fractions are expressed in mass percentage of the total dry weight, after removal of carbonates and organic matter. The reference diameters of the particle classes used in this study are thus: 2, 11, 22, 50, 100, 250, 500 and 1000 µm. The respective cumulative grain-size fractions are indicated as L_2, L_{11}, …, L_{1000}.

Given that the analyses were performed using two different techniques whose interval diameters did not always coincide, intermediate fractions were obtained by linear interpolation of the lognormal distribution to allow both data sets to be combined in one analysis. The average cumulative grain-size distributions were calculated for each stratigraphical unit and are shown in Figure 6.2.

The carbonate and organic matter content was checked for the entire stratigraphical succession, because of the potential importance in determining K_s. Wang *et al.* (2009) demonstrated a negative correlation between hydraulic conductivity and organic matter, which is more important than grain size in the case of a very uniform, well-sorted soil. Hydraulic conductivity falls back to half or maximum one third of its initial value, with an increase in organic matter of 0.5%. This paper deals however with considerably more heterogeneity, both within a single sample, as across samples. Therefore, the role of organic matter is expected to be negligible. Organic matter in the soils occurring at the studied site is on average 6.4% in the first metre below surface (Soetens, 2008). However, the geological sediments subject of this study are mainly part of the first 50 metres below surface. Organic material is hardly present (mostly below 0.3 %, never exceeding 2%). The same is true for the carbonate content within the upper aquifer (below 0.5%), but for the lower aquifer, the average value amounts 3 to 4 %, with an

occasional peak between 5 and 10%. Both the logs obtained from the two geotechnical boreholes are given in Figure 6.3. Given these low percentages, it is assumed that the organic matter and carbonate content only have very little influence on the hydraulic conductivity in comparison with the textural information. The few samples with carbonate content exceeding 5% might however be influenced significantly, and not accounting for this quantity might indeed induce errors. The lack of data for the entire dataset did however not permit the inclusion of this parameter in the analysis, but the presented method can always be extended to include additional parameters.

6.2.1.3 Saturated hydraulic conductivity measurements

In total 308 hydraulic conductivity measurements were performed on cylindrical plugs of 100 cm³ using a constant head permeameter based on the principles of Klute (1965). This was assumed to be more practical than a falling head method, since the amount of samples to be analysed was 173, and the fixed head method allows for a more easy laboratory setup where several samples are measured at the same time. Given the hydraulic conductivities of the studied sediments, these fixed head permeameter tests would have required a few tens of minutes up to a few days. Because of the very low gradients applied, the permeability of the samples is not believed to be altered. An additional 27 hydraulic conductivity measurements were performed using a modified constant head permeameter adapted to medium to low K_s-values as described in Wemaere *et al.* (2003). For the latter, cylindrical samples are transferred into a stainless steel cylindrical cell with two sintered stainless steel filters (pore diameter of 10 µm) at both ends. De-ionised water is injected at the bottom of the sample under a constant pressure of about 600 kPa and K_s is determined from flow measurements using Darcy's law (1985). The modified constant head permeameter set-up is particularly suitable for clayey sediments with K_s in the range 10^{-13} to 10^{-9} m/s. For high clay contents, the measurement can last for three weeks until steady-state flow is achieved. In the case of more silty or sandy samples, the technique is slightly modified using a 1 m high water column with measurement times usually less than one day. The accuracy is about 8 %. Because we are dealing with Tertiary material, subject to groundwater flow for millions of years, the clay particles are assumed to be relatively immobile under such conditions. Microbial growth is not expected to happen within the timeframe of the measurements.

About half of the samples were taken in horizontal direction, whereas the other half was taken vertically, at about 10 centimetres away from the horizontal sampling locations. The combined sample pairs were studied with statistical testing for the equality of the mean and variance of the obtained K_h and K_v data. On average the ratio of geometric mean K_h and K_v values amounts 2. However, the difference between the two datasets was tested using the statistical *t*-test and F-test. The p-values were respectively 0.104 and 0.076 for the upper aquifer, indicating that the difference of the mean and variance of the K_h and K_v populations is not significant at the 5% level. For the aquitard and the lower aquifer, the *t*-test p-values amounted 0.30 and 0.005, and 0.42 and 0.04 for the F-test.

The difference between the K_h and K_v datasets thus is not statistically significant, except when the lower aquifer dataset is considered separately. There, the presence of a few outliers might be the cause of this, as can be seen in the scatterplots of K_v versus K_h in Figure 6.4. This is also indicated by the coefficient of determination values of 0.46, 0.67 and 0.38 for the upper aquifer, aquitard, and lower aquifer respectively. In any case, it is very clear that the K_v values are not systematically lower than the corresponding K_h values, since 35% of the paired samples show higher K_v values. Moreover, the average observed anisotropy is very small in comparison with the total K variability. The K_h estimates will thus provide indications on K_v but sample-scale anisotropy should be accounted for in certain units (see results Chapter 5).

Grain size analyses were however only performed on the K_h samples. This means that the developed models will in principle be adequate for predicting only K_h data. However, because of the strong indications of isotropy at the sample scale, in this case, the predictions will represent both the K_h and K_v values. A total number of 173 pairs of K_s-grain-size data was therefore available, and is used in the further analysis. The K_s kernel density estimate is shown in Figure 6.5. This is a non-parametric way of estimating the probability density

function of a random variable. One of the advantages of the smooth kernel density estimates compared to discrete histograms is their higher statistical efficiency (Scott 1979). Core samples for which only hydraulic conductivity was measured were used to determine the vertical anisotropy factor (Beerten *et al.* 2010).

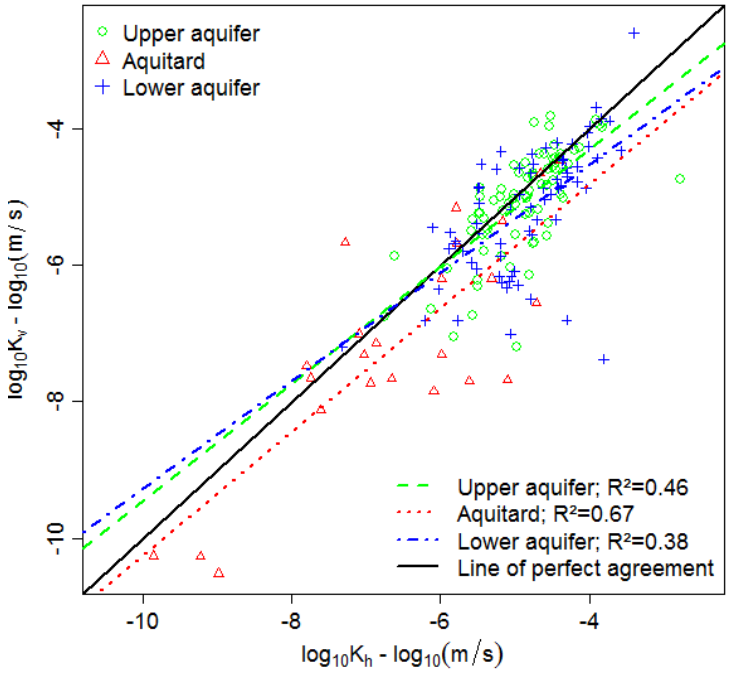

Figure 6.4. Vertical vs horizontal K_s measurements in the three different hydrogeological units, with a linear model fit for each unit.

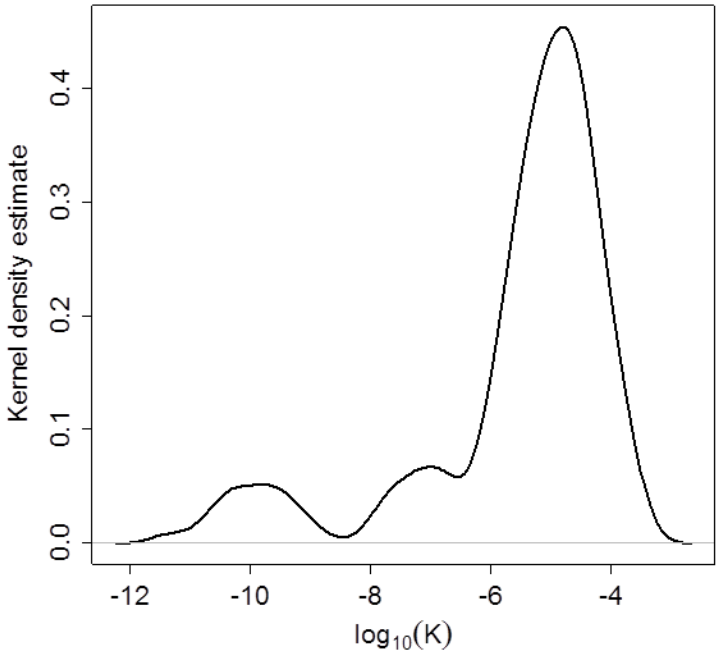

Figure 6.5. Kernel density estimates of all 173 hydraulic conductivity samples used in the model derivation.

6.2.2 Modelling

6.2.2.1 Linear model

All cumulative grain-size fractions (L_2, L_{11}, …, L_{1000}) were used as predictor variables. These are however highly correlated, as indicated in Table 6.1. Regressing on multi-collinear predictor variables has a detrimental effect on regression analysis and has to be avoided (Gunst and Mason 1980). Therefore, multiple regression was applied after orthogonalization of the data using principal components analysis (PCA) as recommended by Van De Genachte *et al.* (1996). PCA produces orthogonal, uncorrelated principal components as new independent variables that are linear combinations of the original variables (Hill and Lewicki 2007). PCA was performed on the covariance matrix, therefore putting more weight on the most variable fractions, because the performance of correlation-based PCA was slightly less. The component loadings and correlation coefficients are given in Table 6.2. By loadings we mean the weight by which each original variable should be multiplied to get the component score. The columns of Table 6.2 equal the eigenvectors of the covariance matrix. The first PC explains 76% of the variance, the first two PCs explain 96% of the variance, and the first three PCs explain 99% of the variance. The largest diameters seem to carry very little information on the variability within the grain-size data, since they only load on a few components. Moreover, they show a very small correlation with K_s, as shown in Table 6.1.

Subsequently, stepwise multiple linear regression (MLR) was performed with the principal component scores. The components with the highest explained variance were added first, and their significance was checked before adding another. All eight PCs were evaluated, even though the higher order PCs accounted for a very small percentage of the variance. This yielded a regression model based on the first (adjusted R^2=0.836), third (0.878), fifth (0.882) and eighth PC (0.888). The contribution of the others PCs in the linear regression is statistically insignificant. The first PC shows negative weights for all grain-size fractions, all being more or less equally important. The third PC has a large positive contribution of L_{100}, while PC5 gets a negative contribution from L_2 and positively from L_{50}. Finally, adding PC8, with contributions from L2, L11, and L22, to the regression model was still statistically significant. A back-transformation is needed to express the regression models in terms of the original particle size parameters. The resulting linear model is presented in Equation 6.3, and some goodness-of-fit measures are presented in Table 6.3.

$$\log_{10} K_s = -3.42 + 1.4x10^{-1}\, L_2 - 3.26x10^{-1}\, L_{11} + 2.76x10^{-1}\, L_{22} - 9.43x10^{-2}\, L_{50} - 1.96x10^{-2}\, L_{100} - 9.23x10^{-3}\, L_{200} - 1.44x10^{-3} L_{500} - 2.04x10^{-3} L_{1000} \quad [6.3]$$

The 95% prediction intervals of this linear model are used to assess the uncertainty about the predicted K_s values (see *e.g.* Kleinbaum *et al*., 2007).

6.2.2.2 GLUE-ANN approach

Artificial neural networks

Artificial neural networks consist, like their biological analogues, of several processing elements, called nodes or neurons, that are interconnected (Figure 6.6). In a feedforward ANN architecture information always moves in one direction; it is the simplest type of ANN as there are no loops in the network. It has input nodes, where the input data are entered, one or multiple so-called hidden layers containing a certain amount of nodes, and one or multiple output nodes, where the target variable predictions are provided.

The number of hidden nodes is a critical parameter of any feedforward ANN: too many nodes may result in overfitting the data resulting in poor generalization on data not used for training. On the other hand, too few hidden nodes will cause underfitting of the model, which will therefore be insufficiently accurate. Most of the mathematics behind the connections and processing within the nodes are in fact linear combinations. The n_1 input variables (x_i^1) are recombined with weights ($w_{i,j}^2$) and a bias (b_i^2) to a set of n_2 new variables (or nodes) in the so-called hidden layer. The outputs of the hidden layers are recombined in a similar way, to provide the networks output. The amount of nodes in the hidden layer is not a fixed value, and depends on the complexity of

Table 6.1. Correlations between grain-size fractions and log_{10}-transformed hydraulic conductivity.

Fraction	ρ (Log10K)	ρ (L_2)	ρ (L_{11})	ρ (L_{22})	ρ (L_{50})	ρ (L_{100})	ρ (L_{200})	ρ (L_{500})	ρ (L_{1000})
L_2	-0.79	1.00	0.99	0.98	0.94	0.85	0.31	0.19	0.12
L_{11}	-0.84	0.99	1.00	1.00	0.97	0.89	0.31	0.18	0.11
L_{22}	-0.84	0.98	1.00	1.00	0.98	0.90	0.31	0.18	0.11
L_{50}	-0.9	0.94	0.97	0.98	1.00	0.96	0.32	0.17	0.11
L_{100}	-0.93	0.85	0.89	0.90	0.96	1.00	0.34	0.17	0.10
L_{200}	-0.43	0.31	0.31	0.31	0.32	0.34	1.00	0.57	0.41
L_{500}	-0.22	0.19	0.18	0.18	0.17	0.17	0.57	1.00	0.84
L_{1000}	-0.14	0.12	0.11	0.11	0.11	0.10	0.41	0.84	1.00

Table 6.2. Variable loadings on the principal components of the grain-size dataset, and the correlation coefficients (in brackets).

Size fraction	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8
L_2	-0.22 (-0.94)	-0.06 (-0.13)	-0.32 (-0.28)	0.00 (0.00)	-0.56 (-0.13)	0.68 (0.08)	0.02 (0.00)	-0.24 (-0.01)
L_{11}	-0.38 (-0.97)	-0.11 (-0.14)	-0.41 (-0.21)	0.01 (0.00)	-0.24 (-0.03)	-0.26 (-0.02)	0.00 (0.00)	0.75 (0.01)
L_{22}	-0.43 (-0.97)	-0.13 (-0.15)	-0.41 (-0.19)	0.02 (0.01)	0.06 (0.01)	-0.51 (-0.03)	0.00 (0.00)	-0.61 (-0.01)
L_{50}	-0.49 (-0.99)	-0.14 (-0.14)	-0.03 (-0.01)	0.02 (0.00)	0.73 (0.08)	0.44 (0.03)	-0.01 (0.00)	0.10 (0.00)
L_{100}	-0.56 (-0.96)	-0.12 (-0.10)	0.75 (-0.26)	-0.04 (-0.01)	-0.31 (-0.03)	-0.10 (-0.01)	0.00 (0.00)	-0.03 (0.00)
L_{200}	-0.25 (-0.45)	0.96 (0.89)	-0.02 (-0.01)	0.12 (0.02)	0.01 (0.00)	0.00 (0.00)	0.01 (0.00)	0.00 (0.00)
L_{500}	-0.03 (-0.25)	0.11 (0.53)	-0.05 (-0.09)	-0.90 (-0.80)	0.02 (0.01)	0.00 (0.00)	-0.41 (-0.10)	0.00 (0.00)
L_{1000}	-0.01 (-0.16)	0.04 (0.39)	-0.02 (-0.07)	-0.41 (-0.78)	0.03 (0.04)	-0.01 (-0.01)	0.91 (0.45)	0.00 (0.00)
Cumulative Variance accounted for (%)	76	96	99	99.68	99.89	99.935	99.997	100

Table 6.3. Performance measures (MSE = mean square error, MAE = mean absolute error, ME = mean error, R^2 = coefficient of determination) for the MLR and GLUE-ANN models. Two indicators for the uncertainty envelope around the ensemble estimate are: M95%PR (= mean 95% prediction interval range in $log_{10}(K)$ space) and 95%PO (= 95% prediction interval outlier fraction).

Modelling all data

Method	MSE	MAE	ME	R^2	M95%PR	95%PO
MLR	0.3	0.42	0.00	0.89	2.23	0.06
GLUE-ANN	0.21	0.34	-0.04	0.93	1.86	0.07

Leave-one-out cross-validation

Method	MSE	MAE	ME	R^2	M95%PR	95%PO
MLR	0.34	0.44	0.00	0.88	2.24	0.07
GLUE-ANN	0.29	0.40	-0.03	0.89	1.80	0.08

the problem (Basheer and Hajmeer 2000). The optimum number of nodes also increases proportionally to the number of observations (Fletcher *et al.*, 1998). The non-linear approximation capabilities of an ANN are a consequence of introducing so-called activation functions in each of the networks nodes. Sigmoid functions are the most commonly used activation functions for the hidden layer, and are applied in this study. For the output node however, a linear activation function was used.

Since the correlations between the predictors and the dependent variable are considerably high (Table 6.1) it is chosen to explain the residuals of the MLR model through non-linear modelling with the ANN approach. Different transformations of the predictor variables were tested for further analysis, but only the one with the best results is retained in this paper, *i.e.* a square root transform. In addition, standardising the inputs can make neural network training faster and reduces the chances of getting stuck in local optima (Sarle 1997). Therefore the cumulative grain-size fractions were standardised to zero mean and unit standard deviation. Furthermore, to avoid difficulties with the training algorithm in performing efficient optimization of the network biases and weights with highly correlated variables (see Table 6.2), orthogonalization by PCA was performed similar to the multiple regression analysis. Because of the universal approximation theorem (Hornik *et al.*, 1989; Hornik 1991) and the continuous character of input and output variables, it was chosen to use the standard feedforward neural network architecture with one hidden layer to develop the model.

The adaptive gradient descent with momentum training method (Tollenaere 1990; Sarkar 1995; Haykin 1999), also known as SuperSAB, was used to calibrate the network weights and biases. This method is a modification of the standard gradient descent backpropagation algorithm that allows for a faster learning process. It is a combination of the momentum and the adaptive back-propagation methods. The momentum method includes a fraction of the last step in the parameter space when adjusting the network weigths. Equation 6.4 shows how the neuron weight $w^{p}_{i,j}(t)$ for training cycle t is calculated based on the partial derivative of the objective function to the net target neuron input δ^{p+1}_{i} (delta error δ; for a full derivation of the generalised delta rule, the reader is referred to Henseler 1995) times the output of the neuron that has to be weighted y^{p}_{j} times the learning rate η plus the weight of the former step $w^{p}_{i,j}(t-1)$ times the momentum parameter α. This causes the optimization to move faster towards the minimum. However, high gradients around the minimum of the objective function might cause the algorithm to jump across. To solve this problem, the momentum method is combined with an adaptive back-propagation, which allows to adjust the learning rate η according to changes of the sign of the objective function gradients. The learning rate η is increased if the sign of the gradient remains the same or decreased if it changes during training by respectively the factors μ or d, and is calculated with Equation 6.5. The initial learning rate η and the momentum parameter α were selected by trial and error, and are equal to respectively 0.01 and 0.5. The same method was applied for the network biases, and the least mean square error (MSE) criterion was used to evaluate model performance during training.

$$\Delta w_{ij}^{p}(t) = -\eta \delta_{i}^{p+1} y_{j}^{p} + \alpha \Delta w_{ij}^{p}(t-1) \qquad [6.4]$$

$$\eta_{i}^{p}(t) = \begin{cases} \mu\, \eta_{i}^{p}(t-1) \; if \; \Delta w_{i}^{p}(t) \Delta w_{i}^{p}(t-1) \geq 0 \\ d\eta_{i}^{p}(t-1) \; if \; \Delta w_{i}^{p}(t) \Delta w_{i}^{p}(t-1) < 0 \end{cases} \qquad [6.5]$$

To ensure generalization capacity of the trained neural network, the early stopping technique was invoked (Prechelt 1998). The network performance is checked regularly on a validation set during training. When the model performance starts to decrease, the network is being overtrained, and generalization can not be guaranteed. Hence, the weight configuration with the best performance is selected. Figure 6.7 shows an example of the evolution of the training and early stopping MSE as function of the number of iterations. With the early stopping technique the network weights at about 100 training cycles are chosen in this example because of their maximum generalization capacity (minimum validation MSE). For a higher number of training cycles, the network becomes over-trained, indicated by the increasing validation set MSE. One third of all samples was selected randomly for early

stopping cross-validation. The implementation of the discussed ANN training algorithms by Castejón Limas *et al.* (2010) in the AMORE package for the R language (R Development Core Team 2010) was used in this paper.

The parameter that remains to be determined is n_2, *i.e.* the amount of hidden nodes. Since the optimal number is dependent on the complexity of the problem amongst other things (Basheer and Hajmeer 2000), it is generally determined by trial and error. Several rules of thumb do exist in literature, *e.g.* the number should not exceed twice the number of inputs (Swingler 1996; Berry and Linoff 1997), and approaches are available to perform a stepwise optimization (*e.g.* Fletcher *et al*., 1998). We choose to select n_2 randomly with a maximum value of three times n_1, the number of input variables. The early stopping procedure will still prevent a network with a large number of hidden nodes to become overtrained, and hence the uncertainty associated with n_2 will be accounted for.

Uncertainty estimation by GLUE

To obtain uncertainty estimations for the predicted K_s values, a Monte Carlo dataset was generated by considering the following random variations in model parameters and data sets: i) the number of hidden nodes (n_2), ii) the initialization of network weigths ($w_{i,j}^p$) and biases (b_i^p), iii) dimensionality reduction of the input data to a random number of input variables (n_1) after principal component analysis, and iv) the selection of early stopping data. A total of 10000 Monte Carlo realizations was used to achieve convergence of the estimated uncertainty. This gives rise to an ensemble of 10000 ANN models, which are all a priori equally likely. To update the likelihoods of the individual models, the concept of GLUE (Beven and Freer 2001) was used.

GLUE is based on a rejection of the concept of the optimal model. Instead, multiple models are considered to be equally acceptable if their performance is alike. This concept is also called equifinality. For the application of the GLUE method to this model ensemble, a number of choices have to be made explicitly (Beven and Freer 2001). First, a quantitative measure of model performance has to be used to assess the acceptability of each model and to weigh each model accordingly in the model ensemble prediction and uncertainty estimates.

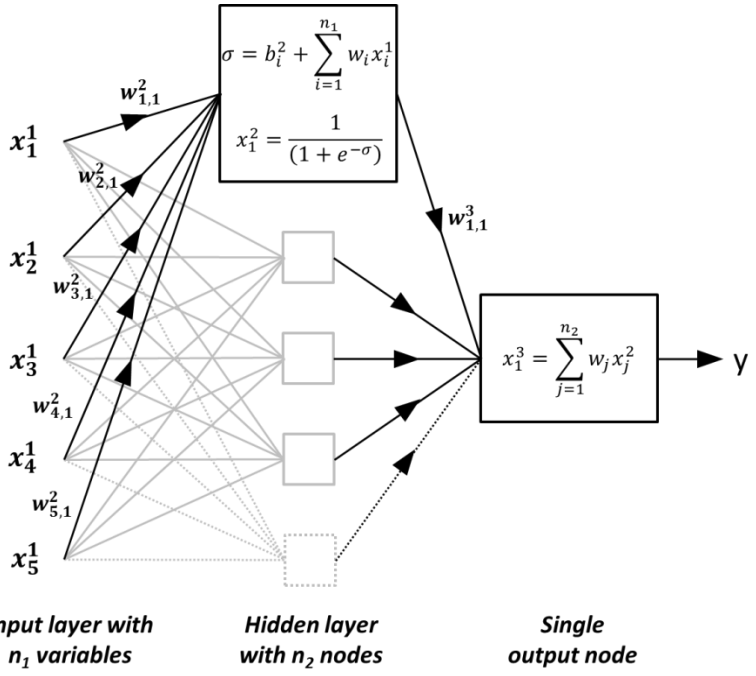

Figure 6.6. Neural network architecture, weights, and activation function.

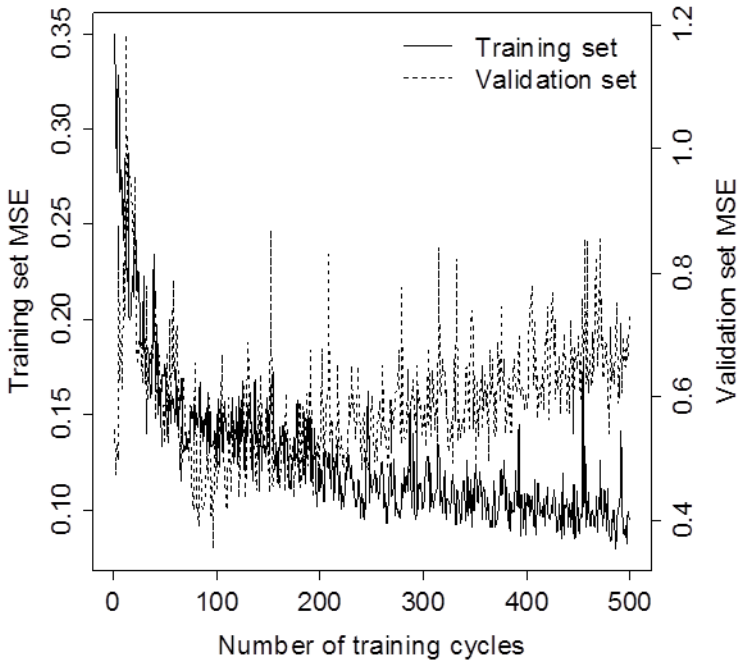

Figure 6.7. Behaviour of ANN model performance (mean square error) as function of training cycles. When the MSE no longer decreases an optimal number of training cycles has been reached, and early stopping is recommended: at about 100 training cycles the maximum generalization capacity is reached. For larger training cycles, the network starts to become over-trained and the validation error rises again.

This measure is called the likelihood measure. A second choice involves the rejection of the lesser performing part of the model ensemble by selecting a behavioural threshold. In a final step, the model ensemble predictions are calculated by taking the weighted mean, median and 2.5 – 97.5 % weighted quantiles from the ensemble of behavioural model predictions, with the individual model likelihoods $L(\theta_i|Y)$ used as weights, possibly rescaled by the behavioural threshold. Considerable controversy exists however about the use of such subjective likelihood measures, but examples exist that are consistent with classical and Bayesian statistics. The approach proposed by Stedinger *et al.*, (2008) was adopted here, which is valid for the assumption of normal and independently distributed model errors. Equation 6.6 is used as the likelihood measure, which equals the model weights,

$$L(\theta_i|Y) = \kappa exp\left[-n\sigma_{\varepsilon,i}^2/2\sigma_\varepsilon'^2\right] \qquad [6.6]$$

where $L(\theta_i|Y)$ is the likelihood for the ith parameter set, θ_i, conditioned on the observations Y, $\sigma_{\varepsilon,i}^2$ is the mean square error for the model with the set of parameters θ_i and σ'^2_ε is the maximum likelihood estimate of the error variance, and κ is a normalising constant. This ensures that i) the best fitting model determines the standard at which other solutions are compared, ii) the probability of each parameter set reflects how well the model matches the calibration data, and iii) the number of samples n used in model derivation is accounted for in determining model performance.

When it comes to applications with the GLUE-ANN approach, *e.g.* for predicting K_s and its uncertainty, the following extension based on Stedinger *et al.* (2008) was implemented. These authors argue that the uncertainty interval of conventional GLUE analysis does not correspond to statistically sound prediction intervals, since the model error that describes the likely difference between the observations and their mean values is generally ignored. Using the likelihood measure in Equation 6.6, one can easily post-process Monte Carlo results to account for model error, and end up with proper prediction intervals. This is accomplished by adding 50 different and independent zero-mean random normal variates with variance $\sigma_\varepsilon'^2$ to each of the model outcomes, as suggested by Stedinger *et al.*

(2008). The number of results in the model ensemble used for constructing 95% prediction intervals thus increased to 50 x 10000. The ensemble of 10000 ANN models was used that way for generating 50 x 10000 predictions for one of the 90-m long boreholes, for which grain-size data was available but not K_s.

6.2.2.3 Model performance assessment

To assess the performance of the developed models, a leave-one-out cross-validation was performed. Due to the limited number of samples, a simplified form of cross-validation where the observations are split into two sets, one for learning and one for control, has not been pursued. For each K_s out of the 173 samples, the model ensemble consisted of 1000 ANN models, trained with the other 172 samples. The number of models was set to 1000 to balance reasonable convergence of the estimated distributions and computational load (two orders of magnitude higher than the normal modelling). For enabling proper comparison with the linear regression, the same leave-one-out cross-validation was applied to the MLR model. The advantage of this approach is that no data should be kept apart to serve as a final test set after model derivation, and hence it is objective. It does require some additional computational time, but it is critical to the success of data-driven modelling methods that rely heavily on the amount of available data, as is the case for ANNs.

Based on the R language, wrapper functions around the AMORE package algorithms (Castejón Limas *et al.*, 2010) for a GLUE-ANN analysis are provided in Appendix A. Several options are provided, including the leave-one-out cross-validation test using different likelihood measures, and a pre-processing routine for the input data. A small script is added to illustrate the use of these functions with a simple example.

6.3. Results

6.3.1 Predictive power for K_s models

First, the performance of the MLR model (Eq. 6.3) and the optimized GLUE-ANN model are compared. The prediction versus observation scatterplots of the model ensemble based on all available samples and the model testing by leave-one-out cross-validation are given in respectively Figure 6.8 and Figure 6.9. Overall, both models perform reasonably well, with all predictions within two orders of magnitude of the observations, and 90% within one order of magnitude. The predictions of the GLUE-ANN model in the K_s range of 10^{-4} to 10^{-6} m/s seem to be less biased. Based on the performance measures provided in Table 6.3, the GLUE-ANN is demonstrated to perform better than the MLR according to all measures, except for the mean error. The difference in performance between both models is however small.

Figure 6.10 shows the prior and posterior distributions of the number of hidden nodes in the model ensemble. As described before, the number was chosen randomly between one and three times the number of variables used as input for the network, and thus has a uniform prior distribution. Comparing this with the posterior distribution, given by applying the weights based on Stedinger *et al.* (2008), it becomes apparent that the optimal number of hidden nodes is 5 (which is also the architecture for the model with maximum likelihood). Networks with more than 12 hidden nodes seem to perform worse, and are thus penalised in the behavioural weighting. Some contribution comes from the models with about 9 hidden nodes.

To demonstrate the superior behaviour of predictive K_s models based on using multiple grain size variables or the entire grain size distribution compared to single parameter models, linear regression was also performed with the highest correlated grain size fraction, L_{100}. The MSE obtained with this model was 0.38, and increased to 0.48 with the leave-one-out cross-validation. Comparing these values with the model performances in case of MLR (Table 6.4) indicates that using a single grain size fraction is 19% less accurate than using the entire grain size distribution (based on the ratio of the root MSE). The difference in accuracy with the non-linear ANN model increases to about 29%.

Next, Table 6.4 provides an overview of the performance of several literature based models to estimate K_s that are discussed by Vienken and Dietrich (2011), and the predictive K_s model of the Rosetta software by Schaap *et al.* (2001). A comparison between these values and the ones for the MLR and GLUE-ANN model indicate the benefit of using a site-specific calibration. Most of them perform worse on the MSE criterion than on the R^2 value, indicating that

relative hydraulic conductivities can be obtained using these literature methods, but correct absolute values are more difficult to obtain without calibration. For example, based on the MSE, the UBSR performs four times worse than the ANN model, while the Beyer model is even more than 40 times worse.

6.3.2 GLUE-based prediction uncertainty

To assess the uncertainty about the estimated hydraulic conductivity obtained by both MLR and ANN methods, the 95% prediction intervals are compared between models. The average 95% prediction interval in $\log_{10}(K_s)$-space is included in Table 6.3, as well as the percentage of observations that fall outside the 95% prediction intervals.

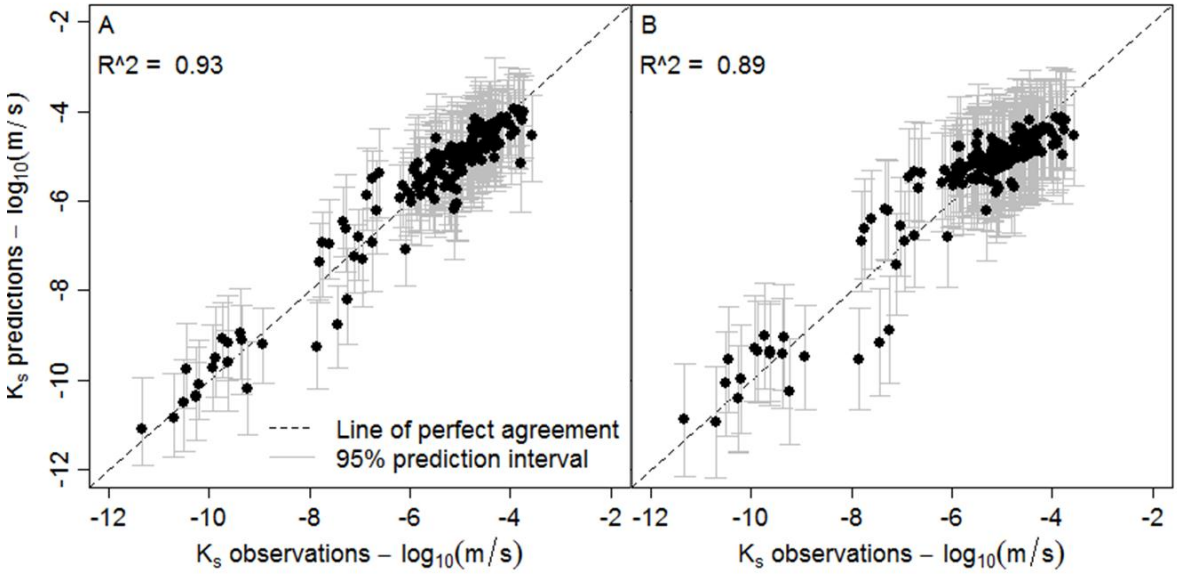

Figure 6.8. Hydraulic conductivity observations versus model predictions and uncertainty estimates (95% prediction interval) for the GLUE-ANN model (A), and the MLR model (B).

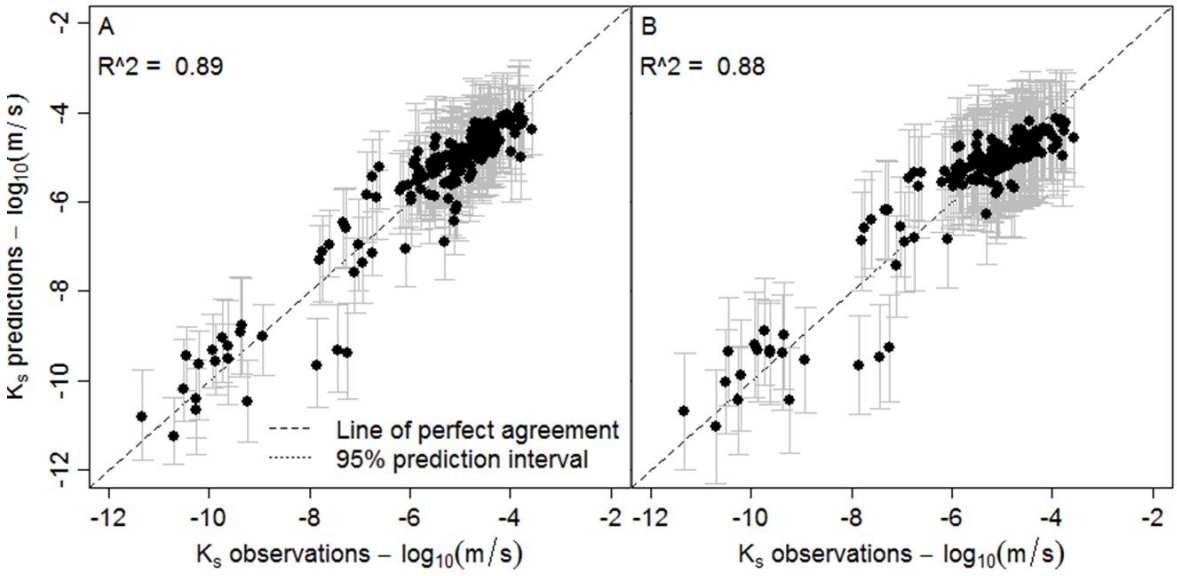

Figure 6.9. Hydraulic conductivity observations versus model predictions and uncertainty estimates (95% prediction interval) from the leave-one-out cross-validation for the GLUE-ANN model (A), and the MLR model (B).

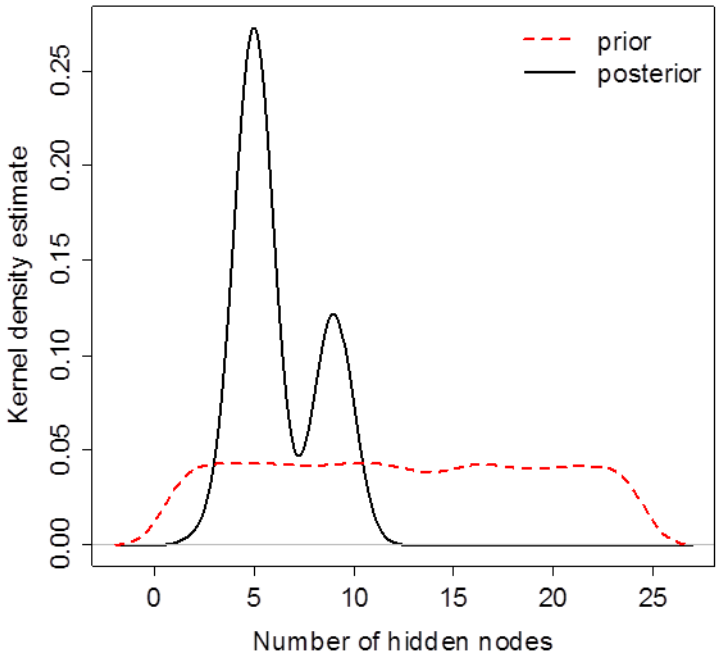

Figure 6.10. Kernel density estimate of the prior and posterior distributions of the number of hidden nodes.

Table 6.4. Performance measures (MSE = mean square error, MAE = mean absolute error, ME = mean error, R^2 = coefficient of determination) for some literature K_s estimation methods discussed by Vienken and Dietrich (2011), and the predictive model of Rosetta (Schaap *et al.*, 2001).

Performance measure	Hazen	Beyer	Kozeny-Köhler	USBR	Seelheim	Kaubisch	Terzaghi	Rosetta
MSE	1.49	9.43	0.89	0.8	2.57	1.71	1.71	2.44
MAE	0.93	2.9	0.74	0.63	1.23	1.05	0.97	1.09
ME	0.52	2.86	-0.42	0.23	1.21	0.99	0.26	0.99
R^2	0.63	0.62	0.73	0.72	0.62	0.71	0.64	0.54

The GLUE-ANN approach appears to reduce the uncertainty about the K_s-values by half an order of magnitude, on average. This is also illustrated in Figure 6.11, by plotting the 95% prediction interval for all samples versus the observed hydraulic conductivity. The mean values are indicated by the horizontal dotted lines. The mean uncertainty owing to the GLUE-ANN model is about half an one order of magnitude smaller than that for the MLR model. Also note that for GLUE-ANN the uncertainty is consistent for the entire range of K_s-values; for MLR the uncertainty increase significantly for $\log_{10}(K_s) > -8$.

To illustrate the very different character of the hydraulic conductivity probability density functions obtained by both models, the distributions for a selected sample are shown in Figure 6.12. The difference between the non-linear and linear methods is very clear. Multiple separate high-density regions can be present in the GLUE-ANN prediction, while the MLR model provides a Gaussian density function. In this particular case, even the third highest peak, which corresponds to the observed value, has a higher density than the MLR estimate, and is thus more likely. Evidently, the MLR model also outperforms the GLUE-ANN model in several other cross-validation cases, but overall, the estimation and uncertainty of the GLUE-ANN model are superior.

6.3.3 Application: predictions of K_s and its uncertainty for a 90-m deep borehole

As illustration of the derived site-specific data-driven multi-parameter models, K_s predictions are made for an additional borehole, for which only grain-size data was available. Figure 6.13 shows the K_s predictions for the entire stratigraphical section.

It seems that the GLUE-ANN K_s predictions are considerably lower in the Kasterlee unit compared to the Mol Lower unit. This is consistent with data from the Dessel 4 borehole (not shown), although the spatial resolution of the latter is only half that of the GT-01 borehole. Since the 90 m long borehole offers at least twice the resolution of the Dessel 4 reconnaissance borehole in terms of grain-size data, the greater detail in predicted K_s supports the use of systematically lower K_s-values in the Kasterlee unit for future groundwater modelling. For the lower aquifer (Diest Clayey Zone and Diest), the predictions of both models seem to be more similar besides some systematic lower values in the most upper and lowest parts. Several layers might now be distinguished in this lower aquifer, since the other boreholes were less deep. For instance, a slightly more conductive layer of about 10 m thickness can be found at a depth of -40 - -50 m above sea level (Figure 6.13).

6.4. Conclusions

It was shown that using grain-size distributions together with linear and non-linear data-driven modelling gives strong improvement over traditional non-site-specific methods for predicting K_s. Multiple linear regression (MLR) combined with principal component analysis is found to be reasonably accurate for predicting hydraulic conductivity on the basis of grain-size data. Up to eight cumulative grain-size fractions were included in the analysis; only two fractions (L_{500} and L_{1000}) did hardly contribute to explaining variability in the K_s data. This shows that nearly the entire particle-size distribution curve contains useful information to be included to build predictive K_s models. This was further demonstrated on the basis of a comparison with a linear regression model using as a single predictor variable the grain size fraction with the highest correlation (L_{100}) with K_s. Considering this single grain size fraction increases the error by 19% and 29% respective to the MLR and GLUE-ANN model performances.

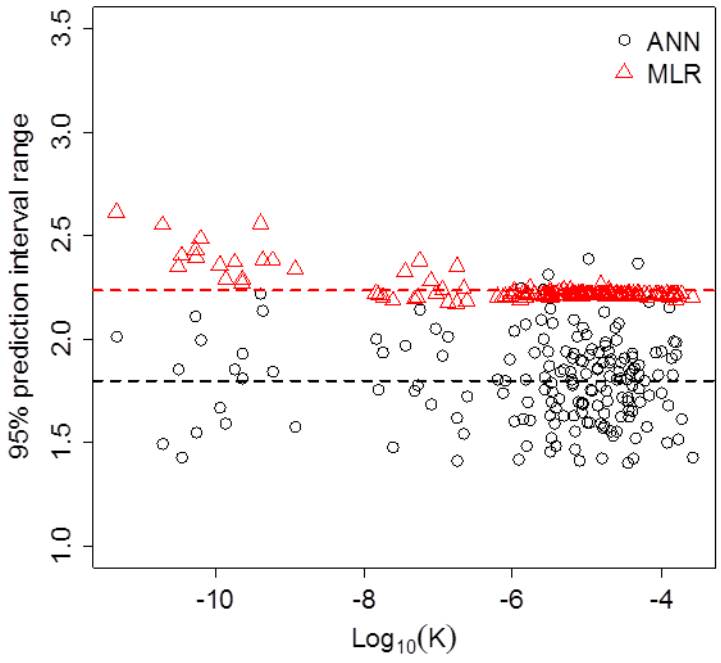

Figure 6.11. Comparison of 95% prediction intervals for the leave-one-out cross-validation.

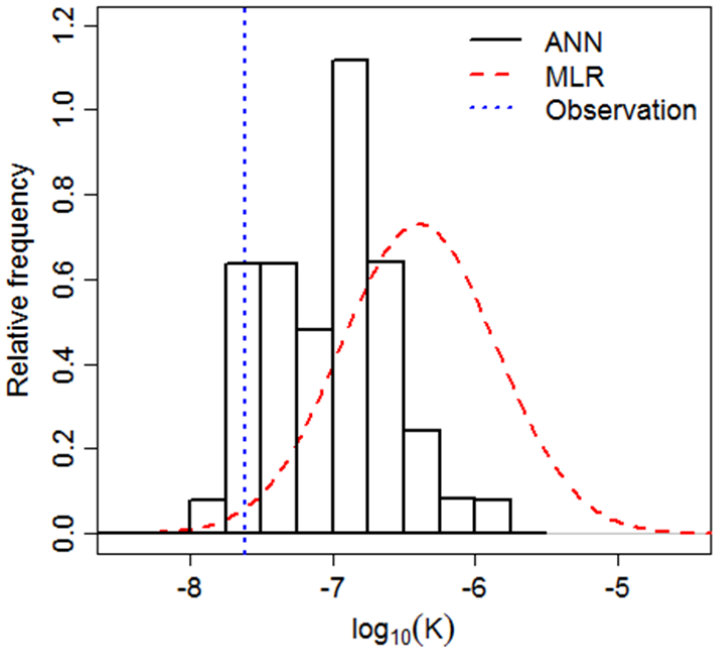

Figure 6.12. Comparison between the predictions of both models for a selected sample from the leave-one-out cross-validation, illustrating the benefit of the non-linear ensemble modelling approach.

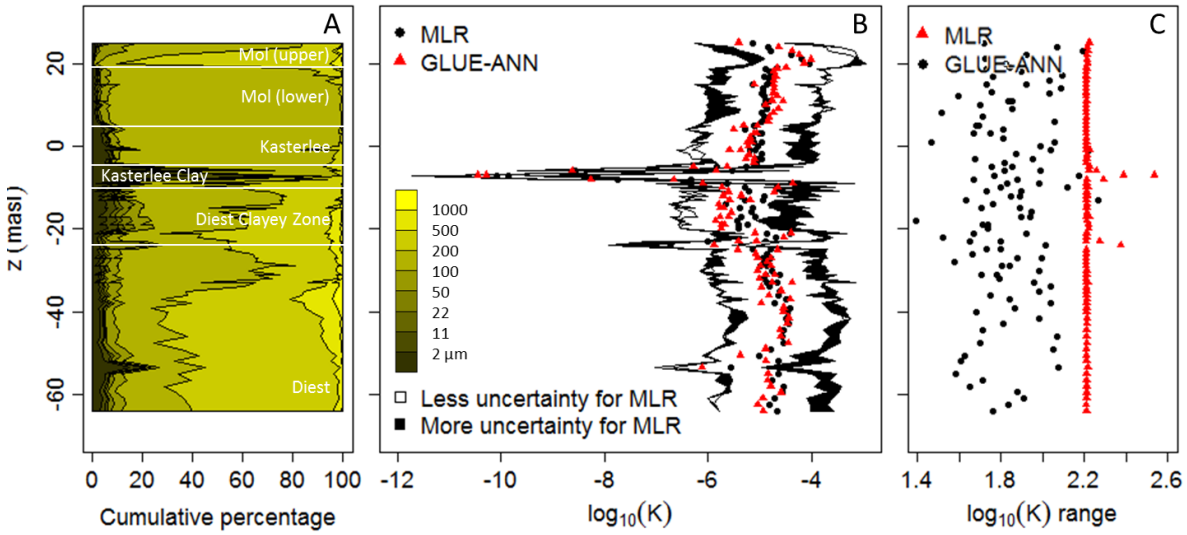

Figure 6.13. A) Grain size data from a 90-m deep borehole (GT 01). B) Application of the GLUE-ANN and multiple linear regression models for K_s predictions using data. The 95% prediction intervals are plotted at both sides of the K_s estimates. If the area in between is colored black, the side closest to the estimate (with the least uncertainty) represents the the GLUE-ANN model. The outer side represents the MLR model. If the area in between has a white color, the MLR model has the least uncertainty. C) Magnitude of the 95% prediction interval for both the MLR and GLUE-ANN models.

The non-linear data-driven modelling was performed by combining artificial neural networks (ANN) and generalized likelihood uncertainty estimation (GLUE). This GLUE-ANN approach provided a greater accuracy in predicted K_s, and considerably smaller prediction intervals, with equal reliability. The GLUE-ANN model is thus preferred over the MLR model, especially in the framework of site specific stochastic groundwater flow modelling where small-scale variability in K_s is important.

To validate the developed models, a leave-one-out cross-validation was performed. The advantage of this approach is that the maximum amount of data remains available for training the ANNs. The GLUE-ANN model proved to be more accurate than the MLR model, and the 95% prediction uncertainty range was reduced on average by half an order of magnitude.

The extension of the proposed model, using additional parameters as organic matter and carbonate content, soil porosity and density could potentially further increase the performance of the models. A hierarchical set of models would then be useful to deal with different data availability (generic, limited site specific, and detailed site specific) at different locations. The results obtained from the textural information only are however already satisfying, and show a reasonable uncertainty. Other missing parameters can be added to the network output layer as well, in order to make predictions for all parameters, at locations with different parameter availability.

The application of the methods to a 90 m long borehole revealed that the increased accuracy of the GLUE-ANN can highlight better the subtle differences between several geological units in terms of K_s. Moreover, the uncertainty associated with the GLUE-ANN estimates is for several stratigraphic units considerably less than the uncertainty on the MLR estimate. The applicability of this site-specific model to other sites remains however limited, although this was not tested. A large range of K_s-values was however used in the model derivation, therefore the application to sites with a similar geological setting may still be successful.

References

Agyare, W.A., Park, S.J., Vlek, P.L., 2007. Artificial Neural Network Estimation of Saturated Hydraulic Conductivity. *Vadose Zone Journal* **6**(2): 423-431.

Asefa, T., 2009. Ensemble Streamflow Forecast: A GLUE-Based Neural Network Approach. JAWRA, *Journal of the American Water Resources Association* **45**(5): 1155-1163.

Basheer, I.A., Hajmeer, M., 2000. Artificial neural networks: fundamentals computing design and application. *Journal of microbiological methods* **43**(1): 3-31.

Beerten, K., Wemaere, I., Gedeon, M., Labat, S., Rogiers, B., Mallants, D., Salah, S., Leterme, B., 2010. Geological hydrogeological and hydrological data for the Dessel disposal site Project near surface disposal of category A waste at Dessel – Version 1 NIROND-TR 2009–05 E, 261.

Berry, M., Linoff, G., 1997. Data mining techniques for marketing sales and customer support. John Wiley and Sons, New York, 464.

Beven, K., Freer, J., 2001. Equifinality data assimilation and uncertainty estimation in mechanistic modelling of complex environmental systems using the GLUE methodology. *Journal of Hydrology* **249**(1-4): 11-29.

Carrier, W.D., 2003. Goodbye Hazen; Hello Kozeny-Carman. *Journal of Geotechnical and Geoenvironmental Engineering* **129**(11): 1054.

Carman, P.C., 1938. The determination of the specific surface of powders. *Journal of the Society of Chemical Industry Transactions* **57**: 225-234.

Carman, P.C., 1956. Flow of gases through porous media. Butterworths Scientific Publications, London.

Castejón Limas, M., Ordieres Meré, J.B., González Marcos, A., de Pisón Ascacibar, F.J.M., Pernía Espinoza, A.V., Alba Elías, F., 2010. AMORE: A MORE flexible neural network package. R package version 0.2-12. http://CRAN.R-project.org/package=AMORE.

Coppola, E.A., Rana, A.J., Poulton, M.M., Szidarovszky, F., Uhl, V.W., 2005. A neural network model for predicting aquifer water level elevations. *Ground water* **43**(2): 231-41.

Cronican, A., Gribb, M., 2004. Literature review: Equations for predicting hydraulic conductivity based on grain-size data. Supplement to Technical Note entitled: Hydraulic conductivity prediction for sandy soils. *Ground Water* **42**(3): 459-464.

Darcy, H., 1856. Les Fontaines Publiques de la Ville de Dijon Dalmont Paris.

Fletcher, L., Katkovnik, V., Steffens, F., Engelbrecht, A., 1998. Optimizing the number of hidden nodes of a feedforward artificial neural network. Neural Networks Proceedings IEEE World Congress on Computational Intelligence. The 1998 IEEE International Joint Conference 2: 1608–1612.

Gunst, F.R., Mason, L.R., 1980. Regression Analysis and its Applications: A Data Oriented Approach. Marcel Dekker, 402.

Haykin, S., 1999. Neural Networks – a Comprehensive Foundation. Prentice Hall, New Jersey, 2nd edition.

Hazen, A., 1892. Some physical properties of sands and gravels. Massachusetts State Board of Health Annual Report 539-556.

Henseler, J., 1995. Back propagation. In: Braspenning PJ, Thuijsman F, Weijters AJMM (eds) Artificial Neural Networks; An Introduction to ANN Theory and Practice. Springer-Verlag, Berlin, 10: 37-66.

Hill, T., Lewicki, P., 2007. STATISTICS: Methods and Applications. StatSoft, Tulsa, OK.

Hornik, K., Stinchcombe, M., White, H., 1989. Multilayer feedforward networks are universal approximators. *Neural Networks* **2**(5): 359–366.

Hornik, K., 1991. Approximation capabilities of multilayer feedforward networks. *Neural Networks* **4**(2): 251-257.

Jain, A., Kumar, A.M., 2007. Hybrid neural network models for hydrologic time series forecasting. *Applied Soft Computing* **7**(2): 585-592.

Joorabchi, A., Zhang, H., Blumenstein, M., 2009. Application of artificial neural networks to groundwater dynamics in coastal aquifers. *Journal of Coastal Research SI 56* (Proceedings of the 10th International Coastal Symposium) 966 – 970.

Khalil, B., Ouarda, T.B.M.J., St-Hilaire, A., 2011. Estimation of water quality characteristics at ungauged sites using artificial neural networks and canonical correlation analysis. *Journal of Hydrology* **405**(3-4): 277-287.

Kişi, Ö., 2007. Streamflow Forecasting Using Different Artificial Neural Network Algorithms. *Journal of Hydrologic Engineering* **12**(5): 532–539.

Kleinbaum, D.G., Kupper, L.L., Muller, K.E., 2007. Applied regression analysis and other multivariable methods. *Cengage Learning*, **906**.

Kozeny, J., 1927. Ueber kapillare Leitung des Wassers im Boden. *Wien Akad Wiss* **136**(2a): 271.

Klute, A., 1965. Laboratory measurements of hydraulic conductivity of saturated soil. In: Methods of soil analysis. Part 1. Black CA et al (eds) *Agronomy* **9**:210-220.

Linderman, M., Liu, J., Qi, J., An, L., Ouyang, Z., Yang, J., Tan, Y., 2004. Using artificial neural networks to map the spatial distribution of understorey bamboo from remote sensing data. *International Journal of Remote Sensing* **25**(9): 1685-1700.

Minasny, B., Hopmans, J.W., Harter, T., Eching, S.O., Tuli, A., Denton, M.A., 2004. Neural networks prediction of soil hydraulic functions for alluvial soils using multistep outflow data. *Soil Science Society of America Journal* **68**(2): 417–429.

Morshed, J., Kaluarachchi, J.J., 1998. Application of artificial neural network and genetic algorithm in flow and transport simulations. *Advances in Water Resources* **22**(2): 145-158.

Nakhaei, M., 2005. Estimating the Saturated Hydraulic Conductivity of Granular Material Using Artificial Neural Network Based on Grain Size Distribution Curve. *Journal of Sciences, Islamic Republic of Iran* **16**(1): 55-62.

Pachepsky, Y.A., Rawls, W.J., Timlin, D.J., 1999. The current status of pedotransfer functions: their accuracy reliability and utility in field- and regional-scale modeling. In: Assessment of non-point source pollution in the vadose zone: Geophysical monograph 108. Corwin, D.L., Loague, K., Ellsworth, T.R. (eds) American Geophysical Union, Washington DC, 223-234.

Prechelt, L., 1998. Automatic early stopping using cross validation: quantifying the criteria. *Neural networks* **11**(4): 761-767.

R Development Core Team, 2010. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

Sarkar, D., 1995. Methods to speed up error back-propagation learning algorithm. *ACM Computing Surveys* **27**(4): 519-542.

Sarle, W.S., (ed) 1997. Neural Network FAQ. Periodic posting to the Usenet newsgroup comp.ai.neural-nets, URL: ftp://ftp.sas.com/pub/neural/FAQ.html, visited on 11/11/2010

Schaap, M.G., Leij, F.J., van Genuchten, M.Th. 1998. Neural network analysis for hierarchical prediction of soil hydraulic properties. *Soil Science Society of America Journal* **62**: 847-855.

Schaap, M., Leij, F.J., 1998. Using neural networks to predict soil water retention and soil hydraulic conductivity. *Soil and Tillage Research* **47**(1-2): 37-42.

Schaap, M., Leij, F., Van Genuchten, M.Th., 2001. Rosetta: a computer program for estimating soil hydraulic parameters with hierarchical pedotransfer functions. *Journal of Hydrology* **251**(3-4): 163–176.

Scott, D., 1979. On optimal and data-based histograms. *Biometrika* **66**: 605–610.

Soetens, T., 2008. Oriënterend bodemonderzoek in het kader van de aankoop van het terrein voor de oppervlakteberging van Umicore. NIRAS/ONDRAF 245.090-MER000, 2008-1119 herz.1.

Stedinger, J.R., Vogel, R.M., Lee, S.U., Batchelder, R., 2008. Appraisal of the generalized likelihood uncertainty estimation (GLUE) method. *Water Resour Res* **44**(12): W00B06.

Swingler, K., 1996. Applying Neural Networks: A Practical Guide. Academic Press, London.

Tiwari, M.K., Chatterjee, C., 2010. Uncertainty assessment and ensemble flood forecasting using bootstrap based artificial neural networks (BANNs). *Journal of Hydrology* **382**(1-4): 20-33.

Tollenaere, T., 1990. SuperSAB fast adaptive Back Propagation with good scaling properties. *Neural Networks* **3** (5): 561-573.

Valverde Ramírez, M.C., Campos Velho, H.F., de Ferreira, N.J., 2005. Artificial neural network technique for rainfall forecasting applied to the Sao Paulo region. *Journal of Hydrology* **301**: 146-162.

Van De Genachte, G., Mallants, D., Ramos, J., Deckers, J.A., Feyen, J., 1996. Estimating Infiltration Parameters From Basic Soil Properties. *Hydrological Processes* **10**(5): 687-701.

Vienken, T., Dietrich, P., 2011. Field evaluation of methods for determining hydraulic conductivity from grain-size data. *Journal of Hydrology* **400**(1-2): 58-71.

Wang, T., Wedin, D., Zlotnik, V.A., 2009. Field evidence of a negative correlation between saturated hydraulic conductivity and soil carbon in a sandy soil. *Water Resources Research* **45**(7). doi:10.1029/2008WR006865

Wemaere, I., Marivoet, J., Labat, S., Beaufays, R., Maes, T., 2002. Mol-1 borehole (April–May 1997): Core manipulations and determination of hydraulic conductivities in the laboratory. SCK•CEN Report R-3590, 56 pp.

Wemaere, I., Marivoet, J., Labat, S., 2008. Hydraulic conductivity variability of the Boom Clay in north-east Belgium based on four core drilled boreholes. *Physics and Chemistry of the Earth* **33**(S1): 24-36.

Zou, R., Lung, W.-S., Wu, J., 2007. An adaptive neural network embedded genetic algorithm approach for inverse water quality modeling. *Water Resour Res* **43**(8): W08427, DOI:10.1029/2006WR005158

Chapter 7

High-resolution hydrostratigraphic characterization of a heterogeneous sedimentary aquifer using cone penetration data

Based on Rogiers B, Mallants D, Batelaan O, Gedeon M, Huysmans M, Dassargues A. High-resolution hydrostratigraphic characterization of a heterogeneous sedimentary aquifer using cone penetration data. Water Resources Research, submitted.

Abstract

Cone penetration testing (CPT) is one of the most efficient and versatile methods currently available for geotechnical and stratigraphical site characterization. Currently available methods for soil behaviour type classification (SBT) of CPT data however have some limitations, and are seldom applied at a regional scale. This paper therefore investigates the use of model-based clustering for SBT classification, and the influence of different approaches on the properties and spatial distribution of the obtained classes. A methodology is proposed for automated lithostratigraphical mapping of regionally occurring sedimentary units, using SBT classification. The methodology is applied to a large CPT dataset, covering ~60 km² at Mol/Dessel, Belgium. Results show that the model-based approach is superior to more frequently applied unsupervised classification approaches or literature classification diagrams, in detecting the true lithological classes. Automated mapping of lithostratigraphical units using advanced SBT classification techniques can yield accurate results, and is more efficient than a time-consuming manual approach.

7.1. Introduction

Cone penetration testing is one of the most efficient and versatile methods currently available for geotechnical and stratigraphical site characterization (Lunne *et al.*, 1997). After being developed at the Dutch Laboratory for Soil Mechanics in Delft (Vermeiden, 1948), its use for soil investigation quickly spread worldwide. While the family of direct push methods has known a great expansion during the last decade (Dietrich & Leven 2009), standard cone penetration tests (CPTs), possibly extended with pore pressure logging (CPTU), are still the most widely used techniques. Due to the maturity of these methods, their speed, costs, precision, accuracy, and repeatability are unmatched today. The classical interpretations of standard CPTs in geotechnical literature are performed by visual examination of the raw data or the use of empirical soil (or soil behaviour type; SBT) classification charts (*e.g.* Robertson *et al.*, 1986; Robertson 1990; Lunne *et al.*, 1997). More recent work in the framework of interpretation or classification of CPT data is mostly focussed on using statistical methods (Cetin & Ozan, 2009), fuzzy classification techniques (Zhang & Tumay 1999; Das & Basudhar 2009), hierarchical and *k*-means clustering (Hegazy & Mayne 2002; Facciorusso & Uzielli 2004; Młynarek *et al.*, 2008; Bilski & Rabarijoely 2009), and the use of neural networks (Bhattacharya & Solomatine 2006; Kurup & Griffin 2006; Kurup *et al.*, 2010), both for supervised and unsupervised problems.

Most of these classification efforts are concentrated on the interpretation of each CPT individually, and classification of CPT data on a regional dataset is mainly limited to use of the classical empirical classification charts (*e.g.*

Rogiers *et al.*, 2010; Schiltz 2008, 2010; Wouters & Schiltz 2012). Moreover, there is a lack of geostatistical interpretations or lithostratigraphical mapping of site-specifically identified SBTs at a regional scale (several tens of km²). Studies of the spatial variability of CPT data are mainly limited to geostatistical treatment of only the vertical direction (Fenton 1999; Uzielli *et al.*, 2005; Kulatilake & Um 2003), 2D interpolation of continuous parameters derived for each single CPT test (Flach *et al.*, 2005; Lenz & Baise 2007), or 3D variography of the raw CPT data or derived continuous parameters (Jaksa *et al.*, 1993, 1997; Tillmann *et al.*, 2008; Liu & Chen 2010). Quantification of spatial variability of SBT classes, indicating the existence of different sedimentary facies or lithostratigraphical units, can however be used at larger scales, and should therefore be investigated.

The above-mentioned unsupervised heuristic clustering methods (hierarchical and *k*-means) show several limitations due to the underlying probability models (Fraley & Raftery 2002). The standard *k*-means clustering algorithm for instance yields equal-volume hyperspherical clusters, which might lead to unnecessary partitioning of the true classes within the data due to a wrong underlying probability model. Moreover, the standard *k*-means algorithm requires that the number of clusters is provided as input, which often is an arbitrary choice. Extensions of the *k*-means algorithm were however developed to overcome this problem. The *x*-means approach (Pelleg & Moore 2000; Ishioka 2000, 2005) is one solution, where a more efficient algorithm is combined with the use of the Bayesian Information Criterion (BIC) to provide both the number of clusters and their parameters. The model-based clustering approach of Fraley & Raftery (2002; 2009) goes even further by using mixture models with the expectation-maximization (EM) algorithm, and generalizes to several underlying probability models. We therefore compare the *x*-means and traditional methods from literature to the model-based clustering approach. Since the use of discrete SBTs is often more robust and facilitates lithostratigraphical mapping, a methodology is also presented for the automated lithostratigraphic mapping of sedimentary units at the regional scale, making use of the site-specific SBT classification, and is compared to the more traditional manual approach.

The clustering algorithms and lithostratigraphical mapping are applied to the CPT dataset of the nuclear zone of Mol/Dessel (Belgium), and assessed with available borehole data (Beerten *et al.*, 2010), lithostratigraphical mapping using a more traditional approach (Rogiers *et al.*, 2010; Schiltz 2008, 2010; Wouters & Schiltz 2012), and the resulting spatial indicator variability. Both the CPT and borehole datasets were gathered by ONDRAF/NIRAS (Belgian Agency for Radioactive Waste and Enriched Fissile Materials) in the framework of the cAt project, *i.e.*, a surface disposal project for low and intermediate short-lived radioactive waste (ONDRAF/NIRAS 2010).

The presented methodology is generic, and provides a more robust site-specific SBT classification approach in which the true intrinsic classes of the data are honoured, as well as an automated way to provide sedimentary unit maps fit for use in environmental investigations or models. Moreover, the model-based classification is probabilistic and therefore fit to incorporate in any type of stochastic analysis.

In the methods section, we describe the tested clustering methods, followed by a short introduction to the (geo-)statistical tools used to visualize the SBT properties and spatial distribution. Subsequently we present the proposed lithostratigraphical mapping approach, and a short description of the site characterization and geology of the presented case study. Then, we give and compare the results of the different classifications. Finally, we compare the obtained maps for a certain lithostratigraphical boundary with the results from the more time-consuming manual approach (Schiltz 2008, 2010; Wouters & Schiltz 2012), and we discuss the differences.

7.2. Methods

7.2.1. Soil behaviour type (SBT) classification

7.2.1.1. CPT variables

The parameters used in existing charts or classifications are all derived from the raw CPT data: the cone resistance q_c, and the sleeve friction f_s (Lunne *et al.*, 1997; an overview of all symbols is provided in Table 7.1). The friction ratio represents the ratio between these two variables:

Table 7.1: List of symbols.

Symbol	Explanation
CPT	Cone penetration test
SBT	Soil behaviour type
q_c; q_t;Q_t	Measured, corrected and normalized cone resistance (MPa)
f_s	Sleeve Friction (MPa)
R_f; F_r	Measured and normalized friction ratio
u	Pore water pressure
a	Net area ratio
σ_{v0}; σ'_{v0}	Total and effective in-situ vertical stress
I_c	Soil behaviour type index
z_{strat}	Stratigraphical depth
z_{masl}	Depth in meter above sea level
z_{mbgl}	Depth in meter below ground level

$$R_f = f_s/q_c \times 100\% \quad [7.1]$$

A correction can be applied for the pore pressure in case of CPTu measurements, which results in the corrected cone resistance

$$q_t = q_c + u(1 - a) \quad [7.2]$$

with u the pore pressure and a the net area ratio determined by the characteristics of the used cone. Stress-normalized equivalents of these variables should be used to account for the in-situ vertical stresses: the normalized cone resistance

$$Q_t = (q_t - \sigma_{v0})/\sigma'_{v0} \quad [7.3]$$

and the normalized friction ratio

$$F_r = f_s/(q_t - \sigma_{v0}) \times 100\% \quad [7.4]$$

with σ_{v0} the total overburden pressure, and σ'_{v0} the effective vertical stress. Jefferies and Davies (1993) introduced the SBT index I_c to represent the radius of the concentric circles in the classification diagram of Robertson (1990). The update of that equation by Robertson and Wride (1998) is the following:

$$I_c = ((3.47 - \log Q_t)^2 + (\log F_r + 1.22)^2)^{0.5} \quad [7.5]$$

This variable basically captures the soil type from the raw CPT data, and carries no or little information on the in-situ soil state (consolidation, cementation, or sensitivity), whereas the 2D classification charts do include this. The correction using pore water pressure (Eq. 7.2) is not applied in this study, since most of the data comes from CPTs and the corrections for the small number of CPTu tests proved to be negligible.

7.2.1.2. Classical approaches

Amongst the existing SBT classification methods in literature, the most frequently used ones are those of Robertson *et al.* (1986) and Robertson (1990), shown in Figure 7.1. The latter one uses the normalized CPT variables to account for the increasing overburden pressure with depth, while the former does not. Moreover, the number of SBTs is also different, with especially more classes in the silt-sand range for the first classification. Though an update was provided recently for the 1986 chart to increase the compatibility with the 1990 chart (Robertson 2010), the original version is used in this paper for comparison with the other approaches since its use is more widespread.
Since these diagrams are defined *a priori* and not site-specific in any way, these classifications are purely descriptive, and probably lack the means of finding the true typology of the data. This is illustrated in Figure 7.1 by plotting the measured CPT data on the diagrams. Several groups of data points clearly intersect different regions of the diagram, and would not be classified as one consistent (though heterogeneous) SBT type. This definitely complicates interpretation of stratigraphy or facies. Here, a site-specific classification might yield a solution. Another approach is the use of SBT index I_c ranges to define SBT classes, like the one presented in Table 7.2 and Figure 7.2 (Robertson & Cabal, 2010). This results however in the same problems as the classification diagrams, but in one dimension.

If the class separation seems to be adequate, like between class 4 and 5 in Figure 7.1B or class 5 and 6 in Figure 7.2, there still is the problem of overlap between classes. This means that points near the class boundaries might have a non-negligible probability of not belonging to the class they were assigned to (with maximum probability).

7.2.1.3. *k*- and *x*-means clustering

As stated before, *k*-means clustering is one of the most frequently used unsupervised clustering techniques, mostly due to the straightforward implementation of the algorithm, and limited computational time. Standard *k*-means clustering minimizes within-cluster distances while maximizing between-cluster distances, through minimizing the objective function

$$\sum_{i=1}^{k}\sum_{j=1}^{n}\left\|x_j^i-\mu_i\right\|^2 \qquad [7.6]$$

with k the number of clusters, n the number of data points within each cluster, x_j^i data point j of cluster i, and μ_i the center of cluster i. This is mostly achieved through the following procedure: 1) choose the number of clusters k, 2) initialize k cluster centers randomly within the multivariate data space, 3) classify all data points to the closest cluster (minimum distance to cluster center), 4) recalculate the cluster centers by taking the average or centroid of all data points, and 5) go to step 3 and iterate until convergence is reached (classification does not change). Since the algorithm is heuristic, many initializations may be required to assure finding the global optimum. In practice however, a small number of initializations is usually sufficient (Telgarski & Vattini 2010). Several versions and more efficient adaptations of this algorithm exist (*e.g.* Forgey 1965; MacQueen 1967; Lloyd 1982; Hartigan & Wong 1979).

The initial *x*-means extension of the *k*-means algorithm (Pelleg & Moore 2000) uses splitting of the clusters after each *k*-means iteration to better fit the data according to the Bayesian Information Criterion (BIC) which approximates the hard to evaluate integrated likelihood

$$2\log p(X \mid M) \approx 2\log p\left(X \mid \hat{\theta}, M\right) - v\log(m) = \mathrm{BIC} \qquad [7.7]$$

With X representing the dataset, $\hat{\theta}$ the maximum likelihood estimate of the parameters θ, M the model, *v* the number of parameters and *m* the number of data points

When the BIC does not improve any further by splitting clusters, the optimal number of clusters is reached. The variance around the cluster centers should however also be considered, as a consequence of the progressive division, as proposed by Ishioka (2000). More recently, a merging operation was added to the algorithm, to prevent unsuitable division of clusters due to the splitting order (Ishioka 2005). We use the implementation of this algorithm in the R language, as provided on the author's website (http://www.rd.dnc.ac.jp/~tunenori/xmeans_e.html). In this paper, we only apply *x*-means clustering, as it is superior to the traditional *k*-means approach, and represents the most frequently used deterministic unsupervised classification algorithm.

The maximum number of iterations is set equal to 1000, which was tested prior to the final analysis to ensure convergence and finding the global optimum, and the initial number of clusters is set to two. As differences in the variable variances might present problems, as not all variables would be considered equally important, data standardization is performed prior to the clustering. For a dataset of ~300000 entries, and the variables considered in this paper, the algorithm execution time was between ~10 and ~40 seconds, on a 2.4 GHz CPU.

7.2.1.4. Model-based clustering

Model-based clustering (Fraley and Raftery 2002) basically consists of fitting a mixture of k multivariate normal densities to a multivariate dataset, where the i-th multivariate normal density Φ_i, parameterized by its mean μ_i and covariance matrix Σ_i, is represented by

$$\Phi_i(x \mid \mu_i, \Sigma_i) = \frac{\exp\left\{-\frac{1}{2}(x-\mu_i)^T\Sigma_i^{-1}(x-\mu_i)\right\}}{\sqrt{\det(2\pi\Sigma_i)}} \qquad [7.8]$$

where $x = (x_1, \ldots, x_N)$ for an N-dimensional dataset. Expectation maximization (EM) can be used to obtain the best fit, given the number of clusters k.

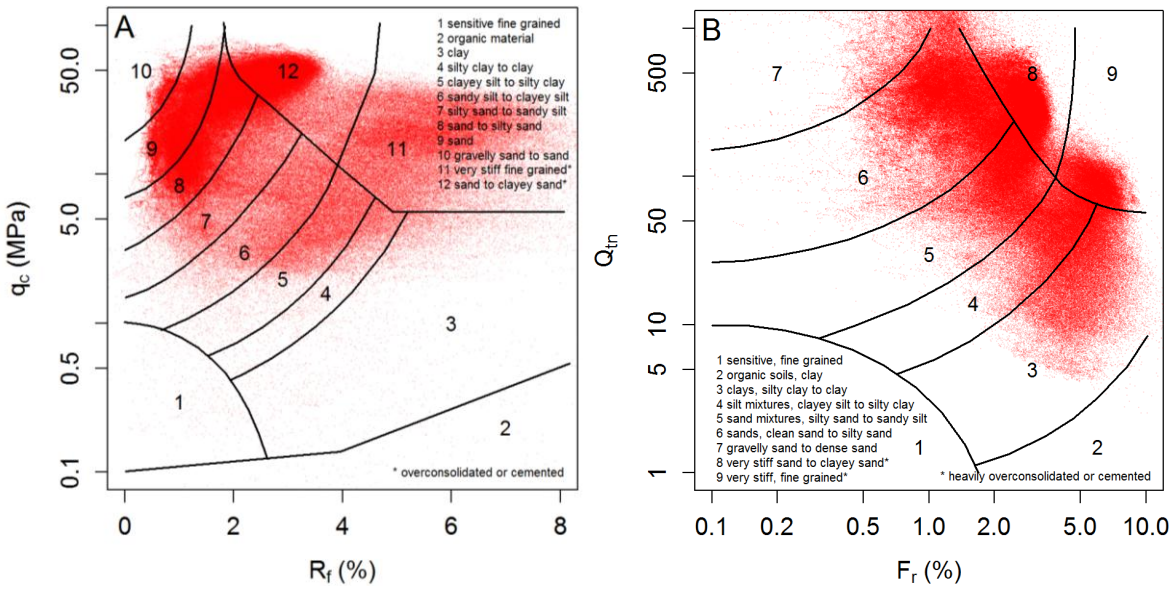

Figure 7.1: SBT classification charts of Robertson *et al.*, 1986 (A) and Robertson, 1990 (B). The data from the case study described further on is shown in both cases as red transparent dots in the background.

Table 7.2: SBT classification based on SBT index (I_c) ranges.

SBT nr.	SBT index (I_c) range	Lithology
1	> 3.60	Organic soils - clay
2	2.95 - 3.60	Clays - silty clay to clay
3	2.60 - 2.95	Silt mixtures - clayey silt to silty clay
4	2.05 - 2.60	Sand mixtures - silty sand to sandy silt
5	1.31 - 2.05	Sands - clean sand to silty sand
6	< 1.31	Gravelly sand to dense sand

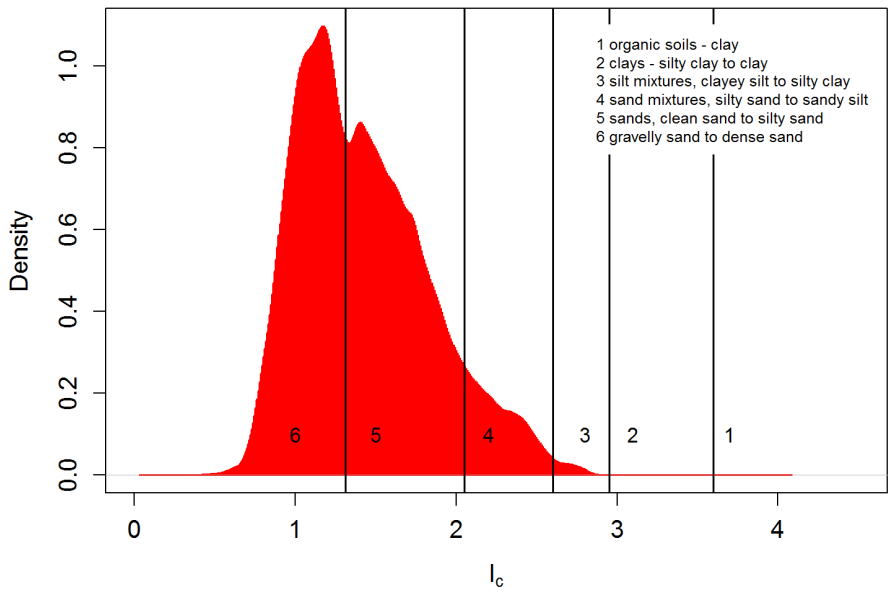

Figure 7.2: SBTs based on the SBT index I_c. The data from the case study described further on is shown as a density plot.

The EM algorithm iterates between two steps: 1) the E step, in which the probability of each observation belonging to each cluster using the current parameter estimates (means and variances) is computed, and 2) the M step, in which model parameters are estimated using the current group membership probabilities.

For the details on the implementation of this algorithm for mixture modelling, the reader is referred to Fraley and Raftery (2002). The software package used is MCLUST (Fraley & Raftery 2009; Fraley *et al.*, 2012). MCLUST can perform model-based clustering for all numbers of classes specified, and a number of different covariance matrix parameterizations. The most simple case is the equal volume spherical model (covariance matrix $\Sigma_i = \lambda I$, with I the identity matrix and λ the common variance), which is similar to the underlying model of *k*-means clustering. The most complicated case is the unconstrained model ($\Sigma_i = \lambda_i D_i A_i D_i^T$, with D an orthogonal matrix that specifies the orientation and A a diagonal matrix that specifies the shape), which allows all clusters to have a different shape, volume and orientation. Hierarchical clustering is used for the initialization of the EM algorithm, and the best model is again selected according to the BIC.

The disadvantage of this algorithm is of course the increase of computational time. For the different cases in this paper, execution times were between ~10 seconds and ~10 minutes. Theoretically, standardization of the variables is not necessary due to the algorithm flexibility, but to avoid problems and speed up the convergence, standardization is applied before clustering, as in the *k*-means approach. The simple *k*-means clustering method is in fact a very specific case of model-based clustering, and the algorithm is a variant of the generalized EM algorithm. The difference between the two algorithms is clearly illustrated in Figure 7.3, which shows the results of clustering a small subset of the CPT data used in this paper, in stress-normalized standard CPT data space. The *k*-means algorithm clearly provides clusters that are more or less equal in size, while the model-based approach does detect the higher density of points in a small area of the parameter space, and assigns a separate cluster to this area. Given this property of model-based clustering, all datasets provided to the algorithms were sampled from the total dataset, with uniform distribution of the stratigraphical depth, z_{strat}. This should avoid biases due to sampling density.

Both *x*-means and model-based clustering algorithms are applied on four combinations of CPT variables: I_c alone; I_c with stratigraphical depth z_{strat}; Q_t with F_r; and Q_t, F_r, and z_{strat}. The reason for including z_{strat} is given in section 7.2.4.

7.2.2. Regional lithostratigraphical mapping

To test the use of site-specific SBT classification in mapping the occurrence of certain lithostratigraphical units or boundaries, we propose the following methodology for automated classification and delineation of such units or boundaries: 1) perform clustering of a selection of CPT parameters with a target amount of 2 classes for the entire dataset, 2) only retain CPTs with a given minimum number of data points of both classes to ensure the boundary actually is penetrated by the CPT, 3) calculate the crossing points of the normalized density estimates for both classes along the z axis, and 4) if the unit that is mapped is relatively small, retain the uppermost crossing point that still has a non-negligible amount of the other class above it, to ensure it is really a transition that is picked. When mapping a unit instead of a lithostratigraphical boundary, the class with maximum density can be assigned to the datapoints.

We apply this approach to the top surface of the aquitard in our case study, which is fairly easy to map manually using traditional SBT classification diagrams as indicators for lithology. The manual mapping of the surface was performed by Schiltz (2008, 2010) and reported in Wouters & Schiltz (2012), and is used as a reference and for validation in this study. For the site-specific clustering, we use both the *x*-means and MCLUST algorithms with I_c as the CPT variable. For the minimum amount of data points in each class we used 150, which corresponds to ~3 m in our case, and the amount of data points above the picked horizon is at least 100 data points; ~2 m. An example of such normalized density estimates and the picked horizon is shown in Figure 7.4.

To create 2D maps of the top surface of the aquitard and the differences with the manual approach, we used the set of retained depths of this horizon, and performed 2D universal

kriging (Goovaerts 1997) to get a continuous representation of the surface.

7.2.3. Visualization of class properties

7.2.3.1. Multivariate statistics

For all of the different obtained classifications in this work, we use biplots to visualize the relationship between the obtained SBT classes and sediment properties available from the cored boreholes, as presented further on. A biplot is an exploratory graph which displays information on both samples and variables of a data matrix (Gower and Hand 1996). We use the first two principal components of the dataset for the axes of the biplot, and project the variables as vectors to this plane using the principal component biplot described by Gabriel (1971). For visualising the different SBT classes, we do not only plot all data samples, but use the cluster centers μ_i. This gives a clear illustration of the relationship between the different SBT classes and the sediment properties obtained from the borehole cores, and correspondingly an idea on the usefulness of the used clustering method. Cluster centers showing the same multivariate properties (plotted near each other in the biplot), therefore might indicate arbitrary division of a single SBT class, while a nice spread of cluster centers indicates that the obtained SBT classes better reflect the true typology of the subsurface sediments. For a more complete description on the construction and interpretation of biplots, the reader is referred to Gower and Hand (1996).

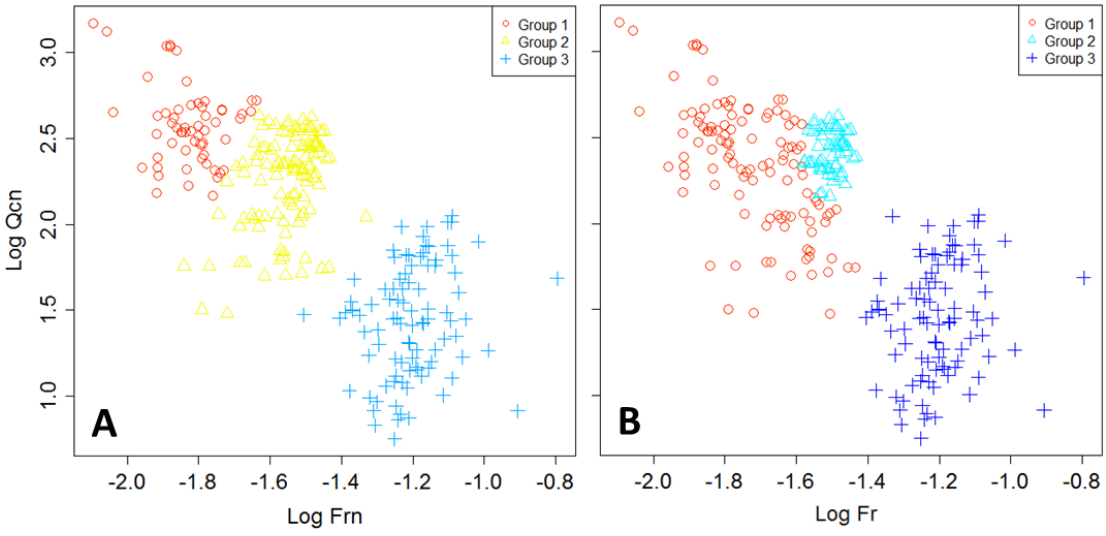

Figure 7.3: A comparison between *k*-means clustering with 3 classes (A), and the model-based clustering (B), for all CPT data where corresponding borehole core analyses are available.

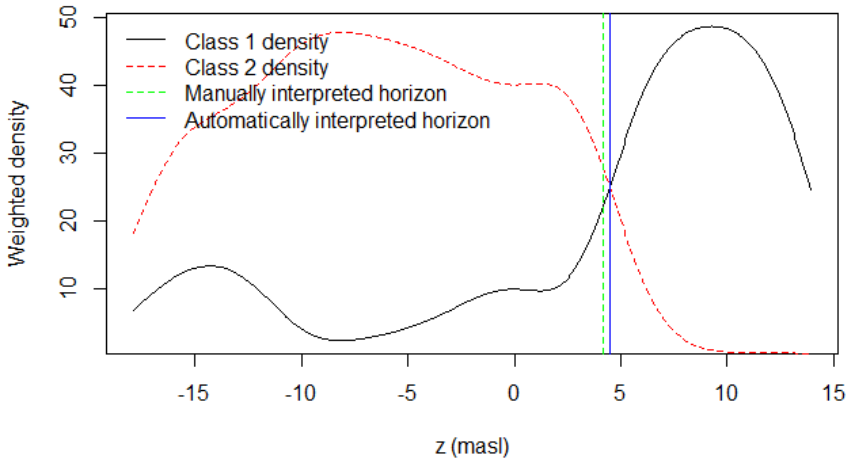

Figure 7.4: example of an automatically picked horizon, mapped by using model-based clustering and the kernel density estimates of the z coordinates of the two contrasting classes.

7.2.3.2. Spatial distribution

To investigate the SBT spatial distribution obtained from the different SBT classification methods, we used the following approach: 1) we determined the marginal distributions of all SBT classes for each meter of the vertical stratigraphical succession, 2) we converted the SBT classifications to *k* SBT indicators, 3) we performed indicator variography using the gstat package (Pebesma, 2004) and fitted spherical variogram models using a least squares approach, with minimum and maximum semi-variance as initial nugget and total sill values and 3 and 1000 m for the initial vertical and horizontal ranges, and 4) we looked at the regional distribution of classes within the entire 3D dataset by using an orthogonal projection of a sideview of the data, perpendicular to the strata dip.

7.2.4. Case study

7.2.4.1. Site characterization

A detailed hydrogeological characterization reaching depths of up to 40 to 50 m (including Quaternary and Neogene formations) has been carried out in 2008-2010 within the Nete basin, at Mol/Dessel, Belgium (Figure 7.5A, B), and was coordinated by ONDRAF/NIRAS (Belgian Agency for Radioactive Waste and Enriched Fissile Materials) within the framework of a surface disposal project for low and intermediate short-lived radioactive waste (ONDRAF/NIRAS 2010). A large amount of quantitative and semiquantitative information has been collected in an area of 60 km², including borehole logs, more than 200 CPTs, and various measurements on undisturbed cores (Figure 7.5C, D; Beerten *et al.*, 2010).

Most of the executed CPTs were standard CPT tests, without the registration of pore pressures, and hence we only work with the standard CPT data. Corrections like Eq. (7.2) were tested for the available CPTu logs, and proved to be negligible. The cone area for all CPT tests was 1500 mm². Reached depths are between 15 and 42 m, with 60% over 30 m.

Several boreholes were drilled at the study area, of which seven were fully cored in the upper 40 to 50 meter. After the drilling operations these cores were sampled, and a range of sediment properties were determined, including saturated hydraulic conductivity, porosity, bulk density, grain size, glauconite content and cation exchange capacity (Beerten *et al.*, 2010).

The CPT tests performed in 2008 were aiming at mapping the presence of an aquitard in the area. In 2010, a smaller area was characterised, and additional CPTs were performed between 1.5 and 5 m from the cored boreholes. This allows to compare the CPT parameters and SBT classification to the derived sediment properties from the borehole cores.

The Quaternary sediments present in the study area mainly consist of different phases of eolian deposited material, or represent alluvial deposits by the larger rivers (see *e.g.* Beerten *et al.*, 2012; Rogiers *et al.*, 2013b; Chapter 3).

The Pliocene Mol Formation consists of white, coarse and medium fine sands. It is sometimes lignitic and can contain some lenses of micaceous clay (Laga *et al.*, 2001), but only the latter has been reported in the current study area. The bottom part is very slightly glauconiferous (less than 2 %) (Beerten *et al.*, 2010). The formation is ~20 m thick in the study area, and is divided into the Mol Upper and Mol Lower Sands (Beerten *et al.*, 2010). The latter are very well sorted and finer and darker in colour, while the former are moderately to well sorted medium sands with a basal gravel layer. Because of the high siliceous content of this sand (99% SiO_2; Sibelco, 2010), it is being mined for various industrial uses (Gullentops and Wouters, 1996).

The Miocene Kasterlee Formation consists of a relatively homogeneous fine, micaceous, slightly glauconitic, sandy upper part (Laga *et al.*, 2001), and a very heterogeneous alternation of clay lenses (clay contents of up to 40%) and sand banks in the lower part (see *e.g.* Beerten *et al.*, 2010; Rogiers *et al.*, 2013a,b; Chapter 2, 3), deposited in a coastal environment with manifest river influence (Louwye *et al.*, 2007). The homogeneous upper part is referred to as Kasterlee Sands while the heterogeneous lower part is named Kasterlee Clay.

The Diest Formation consists of grey green to brownish, mostly coarse and locally clayey glauconiferous sand, often with sandstone layers (Laga *et al.*, 2001). In the current study area, a distinction is made between the clayey upper part of the formation, the Diest Clayey Top, and the Diest Sands below, and glauconite content can be as high as 50 wt% (Beerten *et al.*, 2010).

The geological map including these formations is displayed in Figure 7.6A. The Poederlee (Laga *et al.*, 2001)

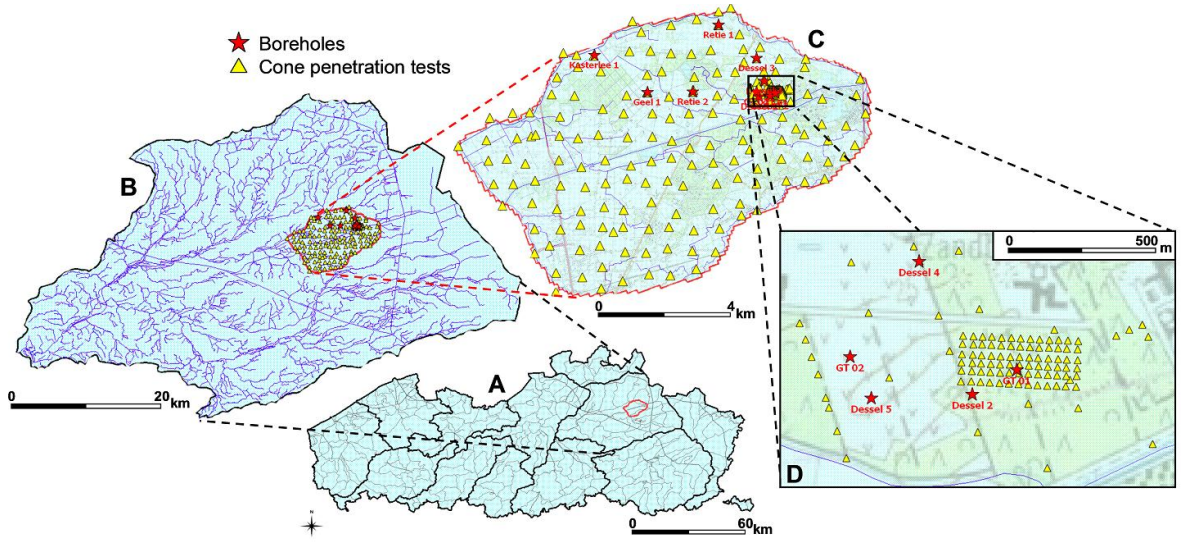

Figure 7.5: Location of the study area within Flanders, Belgium (A), the Nete basin (B). Both CPTs and cored boreholes are indicated on the local groundwater model extent (C), and the disposal site (D).

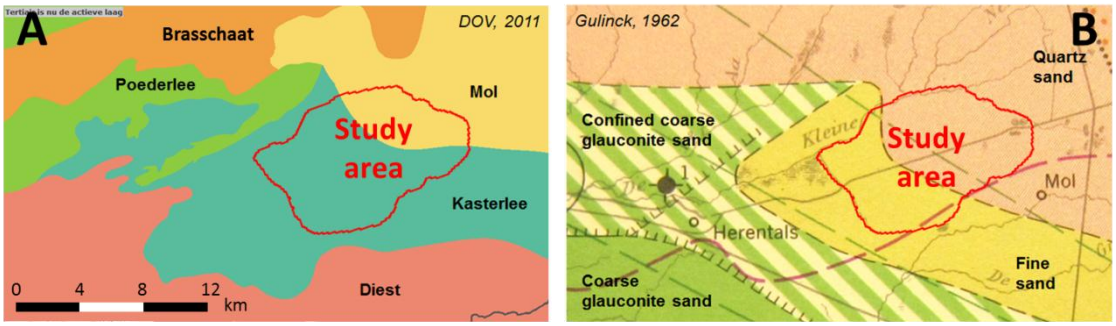

Figure 7.6: Geological (A) and hydrogeological map (B) of the study area and its surroundings, respectively based on DOV (2011) and Gulinck (1962).

and Brasschaat (Gullentops *et al.*, 2001) Formations are lateral equivalents of the Mol Formation, as can be seen on the hydrogeological map (Figure 7.6B), and overlie the Kasterlee Formation (or its lateral equivalents) northeast of the study area. As these lateral transitions are probably more gradual than suggested by the geological map, and we can expect some trends in the sediment properties within our study area. The hydrogeological map also clearly indicates the presence of the Kasterlee Clay aquitard, as part of the coarse glauconite sand (Diest Formation) is indicated as confined.

The local lithostratigraphical succession thus consists of, from top to bottom, various Quaternary deposits, the Mol Upper Sands, the Mol Lower Sands, the Kasterlee Sands, the Kasterlee Clay, the Diest Clayey Top, and the Diest Sands (see Figure 7.8A). Two example CPT logs are displayed in Figure 7.7. The most remarkable features are the high Q_{tn} values for the Quaternary, the low Q_{tn} and high F_{rn} values for the Kasterlee Clay, and less pronounced for the Diest Clayey top, and very different characteristics of the Diest Sands, compared to the Mol or Kasterlee Sands.

While the Quaternary deposits unconformably overlie the Neogene formations, the latter are all inclined, and dipping towards the NE, as shown by the conceptual profiles in Figure 7.8A and the different vertical positions of the layers in Figure 7.7. A sideview of the entire CPT dataset, perpendicular to the layer dip, is shown in Figure 7.8B. This overview clearly shows the differences (both in Q_t and F_r) between the upper aquifer and lower aquifer sediments, separated by the aquitard, which is most clearly

imaged by low Q_t values. The inclination of the layers is also clearly visible.
Schiltz (2008, 2010) has delineated all the lithostratigraphical boundaries, except the Quaternary -Mol Upper and Mol Lower – Kasterlee Sands horizons, within this CPT dataset manually, using the q_c, f_s, and R_f (Eq. 7.1) parameters and the SBT classification by Robertson (1986). The characteristics of these mapped lithostratigraphical boundaries are also reported by Rogiers *et al.* (2010), who used universal kriging to obtain continuous 2D maps.
The top of the Kasterlee Clay aquitard is the most pronounced of these litho-stratigraphical boundaries, and is used in this paper to derive the stratigraphical depth z_{strat}. This is done by substracting the depth of the aquitard of all absolute height values, z_{masl}, so this reference horizon becomes the plane $z_{strat} = 0$. The use of this parameter is tested within the site-specific clustering to investigate the effects on the match between SBT classes and the true lithostratigraphy.

7.3. Results and discussion

7.3.1. SBT classification results

7.3.1.1. Number of classes

The literature SBT classifications were derived for all CPT data, and the *x*-means and MCLUST algorithms were applied to the four datasets, resulting in various numbers of SBT classes, as shown in Table 7.3. The literature classifications show 6, 12 and 9 classes, but not all are present within the dataset as these were not derived in a site-specific way. The *x*-means method clearly yields a smaller number of SBT classes, ranging from 2 to 4 depending on the used dataset. The MCLUST algorithm yields between 4 and 19 classes depending on the used dataset.
This might suggest that the *x*-means approach yields a robust SBT classification as it only divides the data in a few SBTs. However, a too small number of classes might also fail to provide the required amount of detail to discriminate between different lithostratigraphical units. The MCLUST algorithm provides at most 19 SBTs, which at first sight seems too detailed to provide a clearly interpretable overview of the dataset.

7.3.1.2. Multivariate characteristics

The multivariate sediment property characteristics of the different SBT classifications are shown in the biplots in Figure 7.9. The size of the SBT numbers is proportional to the number of data point within the class. To maximize the information gain from the borehole core dataset, missing data at certain depths were completed with values from nearby samples.
Overall, it seems that Q_t correlates positively with z_{strat}, and negatively with F_r and cation exchange capacity, which also relates to the glauconite content. Logarithmic hydraulic conductivity is slightly related to Q_t, and clearly correlates negatively with the clay content of the samples. Porosity and density are slightly related to the latter, but hardly to Q_t or F_r.
The I_c literature classification seems to provide 5 SBTs with clearly distinguishable properties. SBT 6 is however hardly present in the borehole dataset, while SBT 1 is not. Most of the SBT cluster centers fall on the Q_t - z_{strat} – F_r – CEC line, while only SBT 2 clearly deviates from that with a high clay content. This indicates that there is probably clear identification of the aquitard. The literature q_c – R_f classification shows more classes on the same line, with smaller differences in sediment properties, but also several classes with higher clay contents. SBT 3 and 4 almost share the same properties. This might indicate that the SBT delineation might not be adequate for the studied site. The literature SBT classification of Q_t and F_r shows again less classes with well defined sediment properties, but SBT 4 and 9 almost overlap, and might both be an indicator for the Diest Clayey Top.
The *x*-means classifications result in only few SBTs, all along the Q_t – z_{strat} – F_r – CEC line. These classifications might be robust, but clearly lack a separate class for the high-clay-content aquitard, which is the most important feature at this site.
The model-based clustering of I_c seems to deliver the most robust result in which the aquitard is classified separately. The classes most likely represent the Quaternary and Mol Upper Sands, the remaining part of the upper aquifer, the aquitard and the lower aquifer. Including z_{strat} in that classification results in a high number of classes, with a few cluster centers in the upper aquifer data (high z_{strat} values) that overlap. The model-based clustering

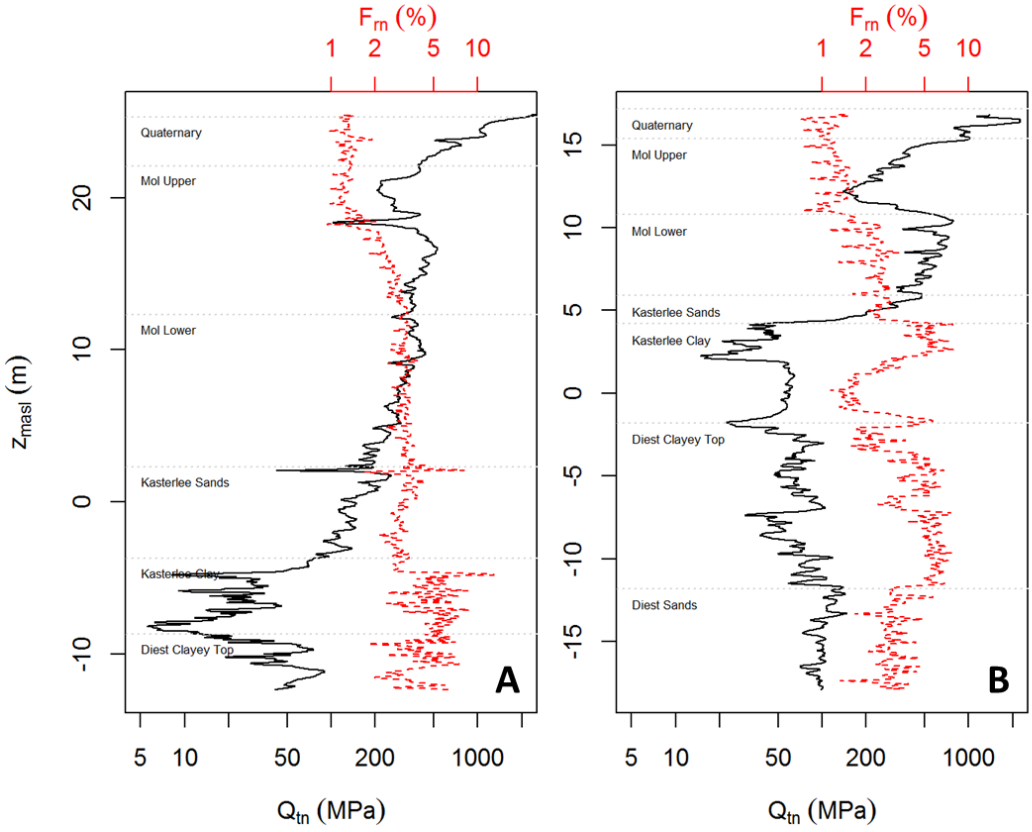

Figure 7.7: Two examples of CPT logs: A) near the Dessel-2 borehole, and B) near the Kasterlee-1 borehole.

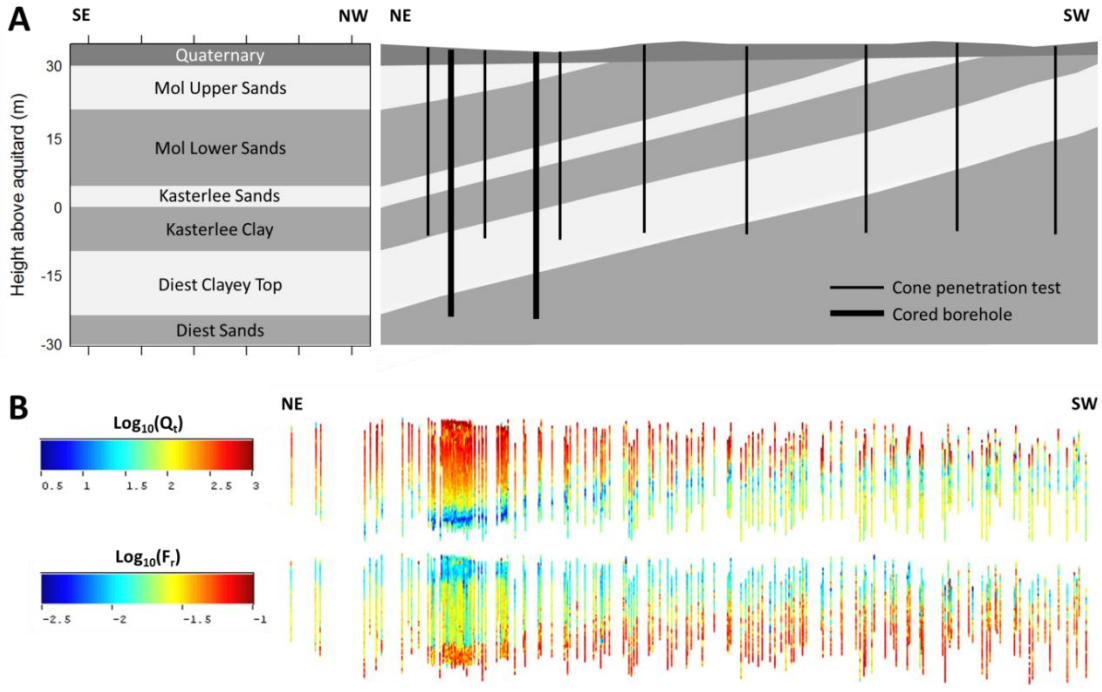

Figure 7.8: A) Conceptual lithostratigraphical profiles through the study area. B) Sideview of the CPT dataset (40x height exaggeration; ~10 km x 40 m) approximately perpendicular to the layer dip, with logarithmic normalized cone resistance (Q_t) and friction ratio (F_r).

of Q_t and F_r shows again some overlap, and detects classes with varying density, porosity, clay content and hydraulic conductivity in the lower part of the lithostratigraphic column. The inclusion of z_{strat} again leads to a larger amount of classes, which, due to the overlap of the cluster centers, seem not to have very different average properties.

7.3.1.3. Spatial distribution of SBT classes

The marginal distribution of all SBT classes along z_{strat} is shown in Figure 7.10, together with the approximate location of the lithostratigraphical boundaries. The different SBT classes are ordered from left to right according to their geometric average Q_t. The top of the Kasterlee Clay aquitard is very clear in all cases, by a large increase in percentage of a single SBT, or the sudden appearance of a new SBT association at a depth of $z_{strat} = 0$. For the *x*-means classifications, the bottom of the aquitard is however not indicated by a decrease of that large SBT percentage, and the aquitard remains undetected. There is thus clearly too little detail within these classifications. Other classifications (literature $Q_t - F_r$; MCLUST I_c) do show a decrease of the low Q_t SBTs, but they remain present within the entire lower aquifer, making delineation of the Diest Clayey Top boundaries difficult. The literature I_c and $q_c - R_f$ classifications clearly show one or more SBTs that correspond to the aquitard, but a distinction between the Diest Clayey Top and the Diest Sands still remains difficult. The three remaining MCLUST classifications do show SBTs that indicate both the aquitard and the Diest Clayey Top, and might be the most useful for use in lithostratigraphical mapping of the stratigraphical boundaries other than the top of the aquitard. Distinction between different units in the upper aquifer is hard to make with the literature I_c and *x*-means I_c and $I_c - z_{strat}$ results. The other literature and *x*-means classifications do provide some indication for the Quaternary en Mol Upper Sands, but the Kasterlee Sands remain difficult to distinguish. The same holds for the MCLUST I_c results, but the other MCLUST classifications, especially those using z_{strat} seem to be more informative. The occurrence of different SBTs are clearly associated with the Quaternary, Mol Upper, Mol Lower and even Kasterlee Sands.

The *x*-means results thus provide too little detail, while the high number of classes in the MCLUST results seem to be very suited for lithostratigraphical mapping using SBT associations, or single SBTs as indicators for lithostratigraphical units.

The same side view of the total CPT dataset as in Figure 7.8B is shown in Figure 7.11, with the resulting SBT classes instead of the continuous CPT parameters. The same observations as from the marginal distributions more or less hold for this data, but a clear difference exists between the literature and site-specific classifications. For the literatures classification, especially $q_c - R_f$ and $Q_t - F_r$, different classes almost randomly alternate at short distances within certain sections of the upper aquifer. This might indicate that the separation between these classes is indeed purely artificial, and in reality we are dealing with a single SBT, that covers multiple sections of the classification diagram. This does not occur with the site-specific classifications, at least not at such a short distance. The MCLUST results including z_{strat} do show lateral variations, but on a more regional scale, which might indicate lateral trends of gradually changing properties, as one would expect from the geological background. The direction along which these changes occur is however not parallel to the layer strike, as we would expect from the known lateral equivalents, but perpendicular to that; parallel to the layer dip, and the Roer Valley Graben faults that have disturbed the studied sediments northeast of the study area (see *e.g.* De Craen *et al.*, 2012).

As the *x*-means I_c results only show two classes, they do not provide a lot of information, but for automatic mapping of the top of the aquitard, this would be an excellent classification.

As we are dealing with 87 SBT indicators for the variography of all SBT results, we provide these in Appendix B, and only discuss a few typical examples of the SBT variograms, presented in Figure 7.12. The first set of horizontal and vertical variograms is from the literature $q_c - R_f$ classification, and shows clearly a pure nugget, or even a hole effect in the horizontal direction. This is due to the splitting of a single lithology type in different SBTs. The resulting random horizontal alternation between SBT 8, 9 and 12 are clearly visible in Figure 7.11. In the vertical direction spatial correlation is clearly present, meaning

Table 7.3: Number of obtained classes for the different SBT classification approaches.

Method	*k*
Literature I_c (Robertson and Cabal 2010)	6
Literature q_c – R_f (Robertson 1986)	12
Literature Q_t – F_r (Robertson 1990)	9
***x*-means: I_c**	2
***x*-means: I_c - z_{strat}**	2
***x*-means: Q_t - F_r**	3
***x*-means: Q_t - F_r - z_{strat}**	4
MCLUST: I_c	4
MCLUST: I_c - z_{strat}	12
MCLUST: Q_t - F_r	14
MCLUST: Q_t - F_r - z_{strat}	19

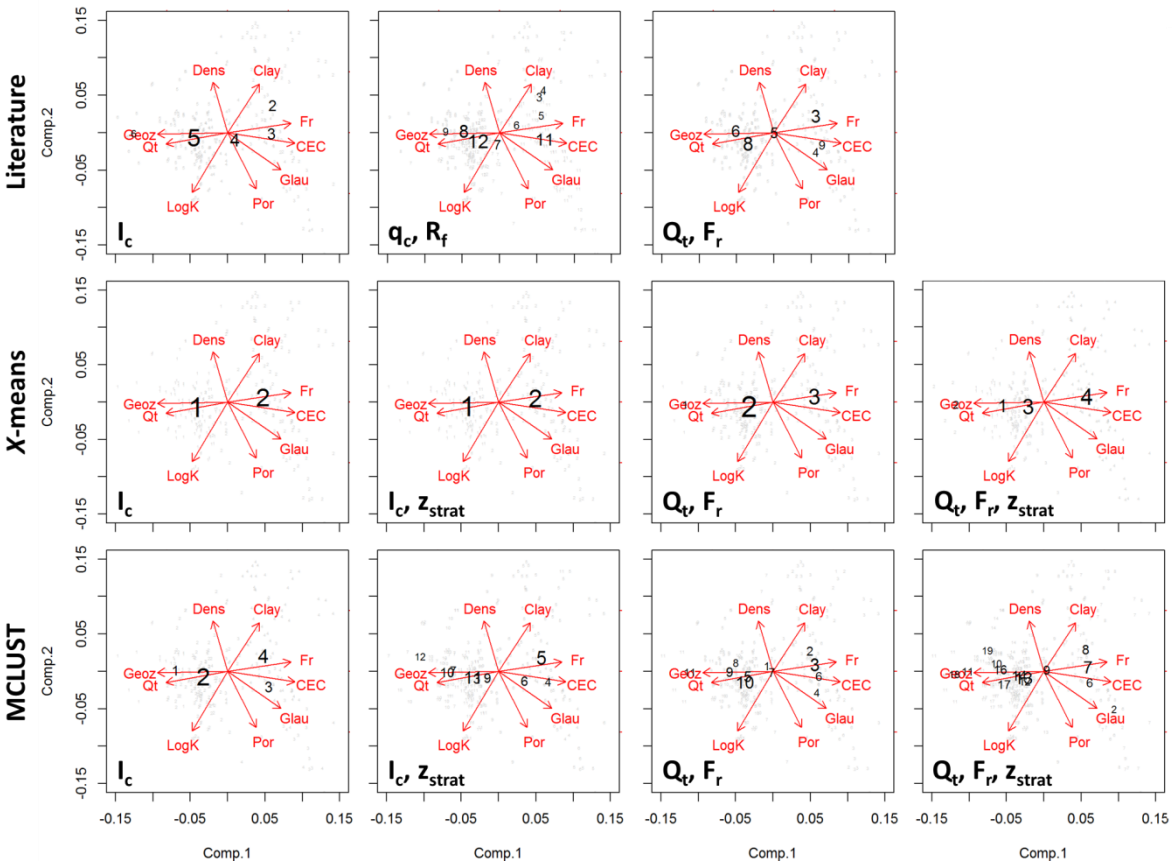

Figure 7.9: Biplots for all 11 SBT classifications. The SBT cluster centers are plotted as black numbers. The x- and y-coordinates of these points are multiplied by 3.5 to illustrate more clearly the relationship with the variables. The size of the numbers is proportional to the amount of data points in the cluster. "Geoz" corresponds to z_{strat}, "Por" to porosity, "Dens" to bulk density, "LogK" to logarithmic hydraulic conductivity, "Clay" to clay content, "Glau" to glauconite content, and "CEC" to cation exchange capacity.

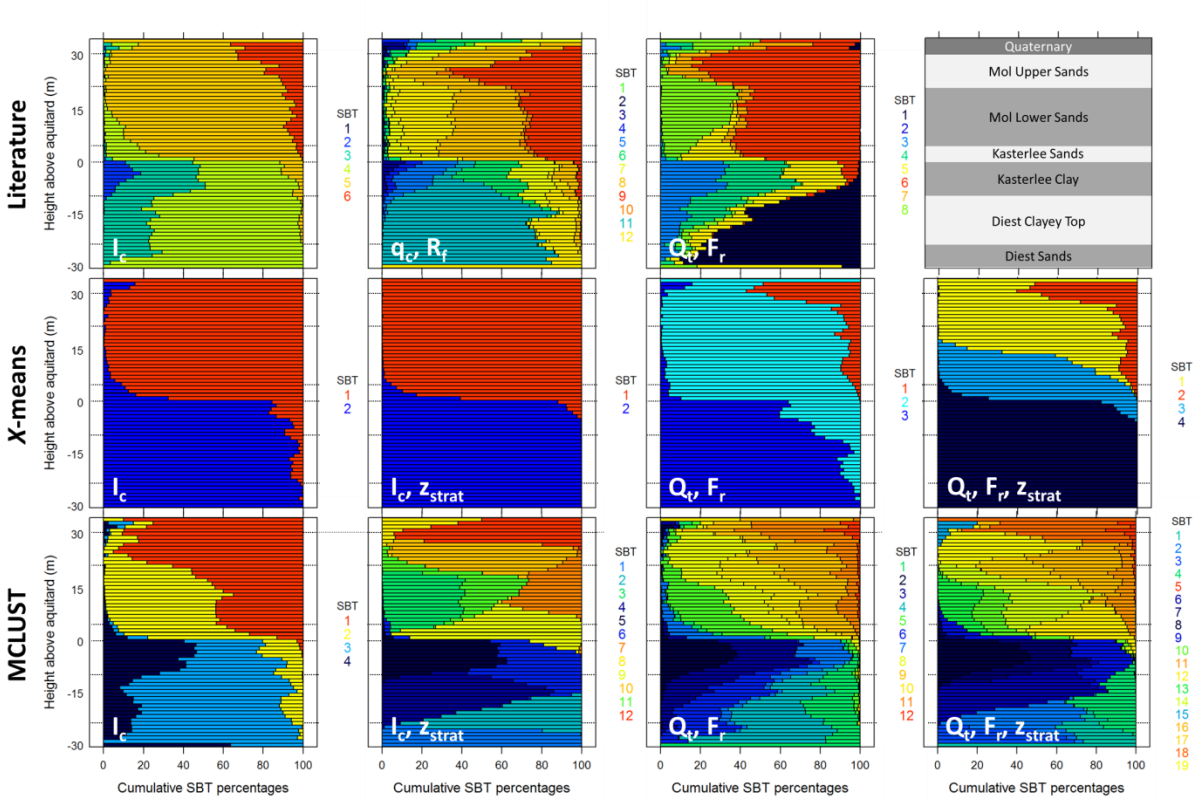

Figure 7.10: Marginal distribution of the different SBT classifications along z_{strat}.

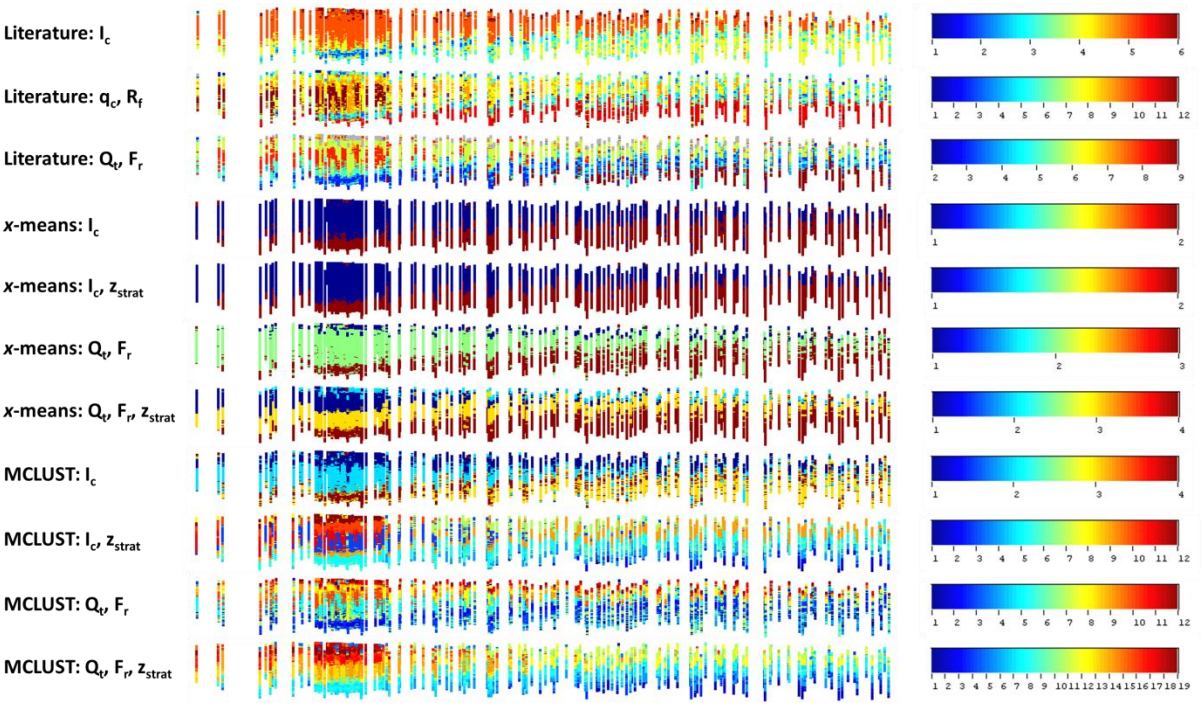

Figure 7.11: Side views of the CPT dataset (20x height exaggeration; ~10 km x 40 m) approximately perpendicular to the layer dip, with SBT classification results for all approaches.

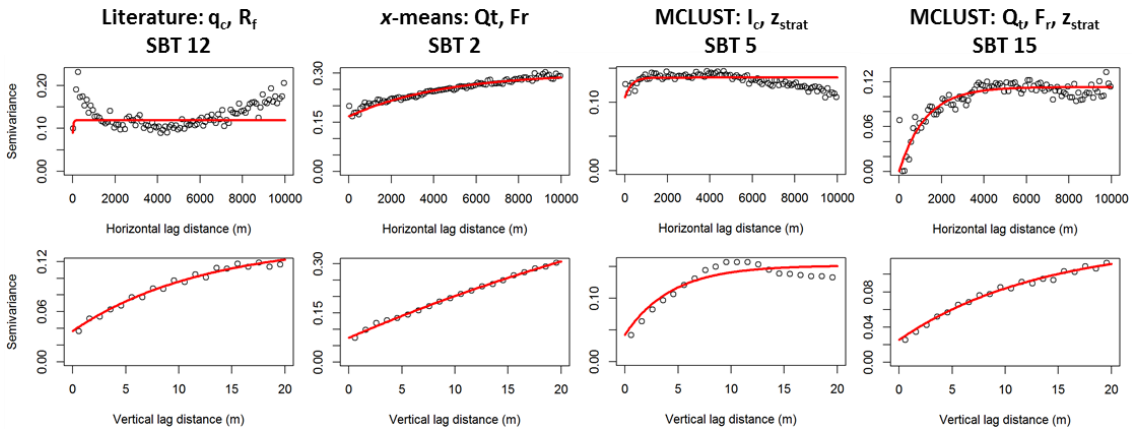

Figure 7.12: Four examples of typical SBT variograms. The full list of variograms is provided in Appendix A.

that there is at least one section in a large part of the CPTs, within the vertical lithostratigraphical column, that is consistently classified as SBT 12. As the distance between two points gets higher, there is an increased chance of transitioning to another SBT. The second example shows an almost linear increase of the semivariance, both in horizontal and vertical direction, though the horizontal variogram starts at a very high nugget. This indicates again that the SBT is located at a specific depth within the lithostratigraphical column, but the linear shape suggests that the maximum thickness of that zone is beyond the largest lag distances considered, so no plateau is reached in the variogram. Figure 7.10 and Figure 7.11 clearly illustrate that this is indeed the case. The horizontal variogram suggests that there is a slight change in proportion of the SBTs in the horizontal direction, and at regional scale. The third example clearly shows a vertical effective range of 10 to 15 m, and almost a pure nugget in the horizontal direction. This indicates a clearly defined section within the lithostratigraphy, no more than ~10 m thick, is classified as SBT 5, and that there is considerable alternation with SBT 6. These classes represent the clayey and sandy parts of the heterogeneous Kasterlee Clay. The final example shows a clear plateau in the semivariance for the horizontal direction. This is an example of the regional-scale lateral changes that are indicated by the SBT classifications, and the horizontal range of ~4000 m is a measure for the horizontal extent of the sedimentary body labelled with SBT 15.

The literature SBT variograms in Appendix A show a mixture of these different types. Pure nuggets or very short ranges occur often in horizontal direction, and the relative nugget values in vertical direction are always high. This indicates that there is considerable random alternation between SBT classes, which is explained by the non-site-specific nature of these classifications. Most *x*-means SBT variograms show a linear increase in the vertical direction, and pure nuggets or only a slight increase in horizontal direction, as in the second example. This indicates the pronounced horizontal and vertical continuity of the *x*-means SBTs, in comparison with all other approaches, and the lack of identification of different lithology types within a single lithostratigraphic unit. The model-based SBT variograms consist of a mixture of different types, similar to the literature SBT variograms, but SBTs with a clear horizontal range occur more frequently, indicating the detection of regional lateral changes of the sediment properties. Also, the relative nugget values are generally lower than for the literature SBT variograms, indicating a higher degree of continuity, hence more robust classification.

The results for the two CPT logs displayed in Figure 7.7 are presented in Figure 7.13. The literature I_c classification clearly does not provide the means to discrimate between the different upper aquifer sands. The difference with the clayey units (Kasterlee Clay and Diest Clayey Top) is however very clear, as is the contrast with the lower aquifer Diest Sands. The literature $q_c - R_f$ and $Q_t - F_r$ classifications show more SBT classes, and allow for more clear identification of the different layers. A clear contrast between the upper and lower aquifer is however not always present, and different SBTs are superfluous as their occurrence corresponds

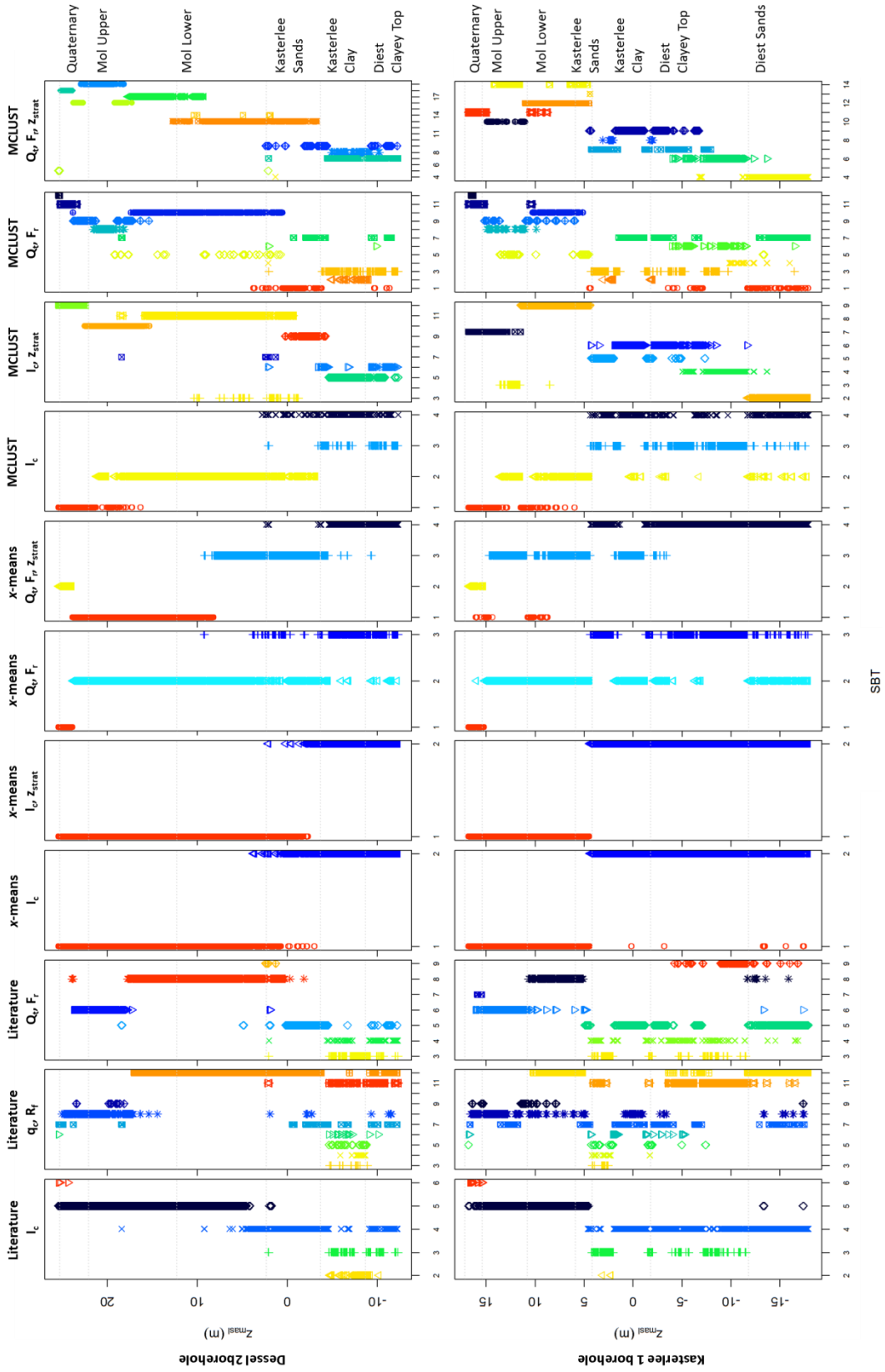

Figure 7.13: Example SBT logs for the CPT data displayed in Figure 7.7.

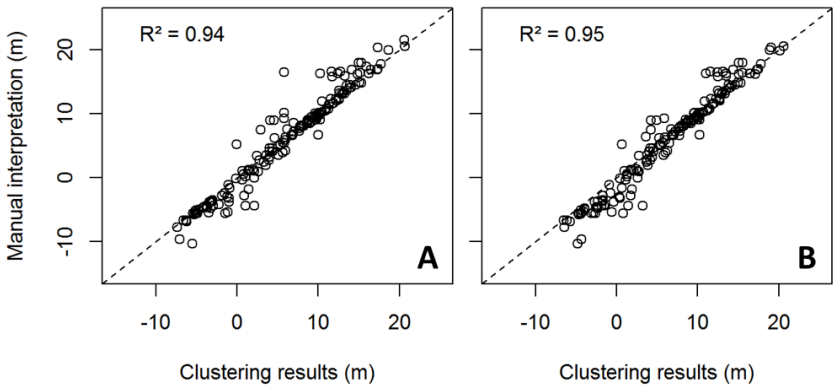

Figure 7.14: Scatterplot of lithostratigraphical mapping of the top of the aquitard versus the manual interpretation of Schiltz (2008, 2010), using A) x-means clustering and B) model-based clustering.

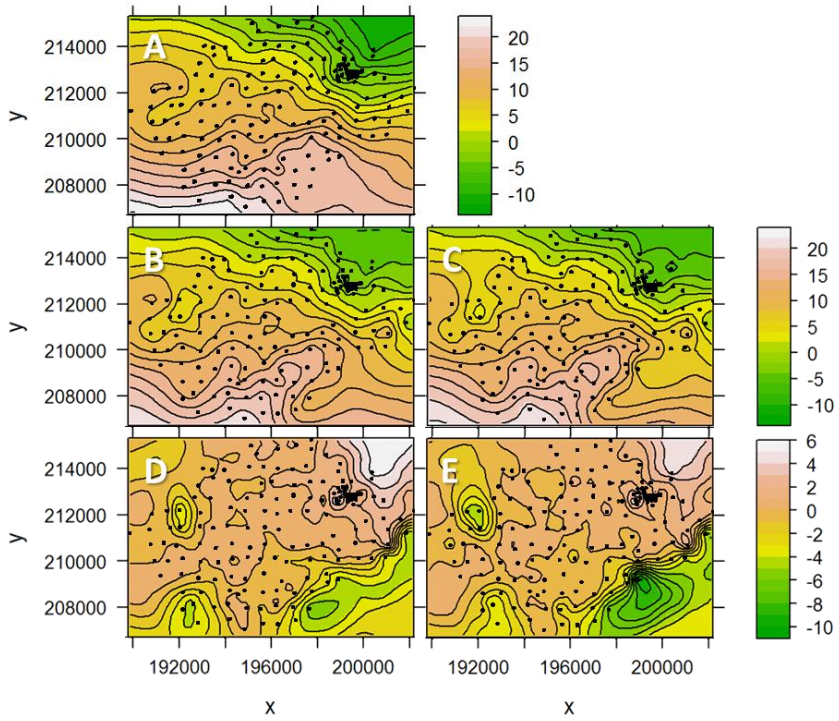

Figure 7.15: Contour maps of the elevation of the top of the aquitard, using A) a manual approach, B) the model-based and C) the x-means clustering. The difference between the automatically derived elevations and the manual derived reference values are presented in D) for the model-based and E) for the x-means clustering. Note that the differences at locations outside of the CPT characterization area are influenced by extrapolation.

exactly to other SBTs (3-5 in the former, and 3-4 in the latter classification). The *x*-means classifications provide very little information, and only succeed in detecting the top of the aquitard, and the Quaternary in some cases. The lower aquifer seems not to be present, as was already clear from Figure 7.10 and Figure 7.11. The MCLUST I_c results provide similar information, but the different units are more clearly identified by including z_{strat}. For the MCLUST $Q_t - F_r$ and $Q_t - F_r - z_{strat}$ classification, most lithostratigraphical boundaries can be linked to the appearance or dissappearance of certain SBTs or SBT associations. The bottom of the Kasterlee Clay would however end up a few meter lower at the Kasterlee-1 borehole, and the Mol Upper – Mol Lower boundary would be interpreted a few meter above the currently interpreted boundary at the Dessel-2 borehole, but the current

interpretation is based on the grain size profile (see Beerten *et al.*, 2010).

7.3.2. Automated lithostratigraphical mapping results

As we are mainly interested in mapping the Tertiary lithostratigraphy, the heterogeneous Quaternary data was discarded prior to the mapping analysis, to avoid inference of this data in the automatic detection of layer boundaries. This was done by removing the upper 3 m of each individual CPT test, as the average depth of the Quaternary in the cored boreholes is ~3 m.

The *x*-means and MCLUST algorithms are applied to the I_c data to obtain two SBT classes. For the *x*-means, this corresponds exactly to the results previously discussed, as the use of 2 SBTs was most optimal according to the BIC. The results of using the kernel density estimates of z_{masl} to pinpoint the top of the aquitard are plotted in Figure 7.14 versus the manual interpretation by Schiltz (2008, 2010). Both approaches show a reasonably good result, with R^2 values of 0.94 and 0.95. The maximum deviation amounts to 10.8 and 7.4 m, but ~60% of the data is within 0.6 and 1 m of the manually identified horizon and 25% within 0.18 and 0.26 m.

The contour maps resulting from universal kriging of the identified horizon locations are shown in Figure 7.15. The main misfit between the manually and automatically interpreted results occurs at the southern border of the study area. In this area, the upper aquifer is only a few meters thick, and hence mapping of the top of the aquitard is more difficult than in the other regions. As the manual interpretation also accounted for identified locations in nearby CPTs, and the general trend, this interpretation is probably more accurate and therefore it was used in the above clustering approaches to obtain z_{strat}, but it remains subjective.

The automatically obtained results seem to suggest the presence of channels in the SW-NE direction, which could be related to the fluvio-marine or estuarine depositional environment.

7.4. Conclusion

In this paper, we have shown that model-based SBT classification of CPT data can be useful for regional lithostratigraphical mapping. The obtained SBTs provide more detailed information thasn those obtained with frequently used deterministic unsupervised clustering algorithms like *k*- and *x*-means clustering. Moreover, the obtained classification honours more the intrinsic classes within the data, in contrast to the classical literature SBT classification charts. These findings are obtained by looking at the multivariate sediment properties from cored boreholes, in function of the SBTs, or by studying the spatial distribution of the obtained classes. The use of stratigraphical depth for clustering proved to be useful for the presented case study, and is recommended for geologically layered sites (unconsolidated sedimentary rocks).

We also proposed a methodology for automated lithostratigraphical mapping using site-specific SBTs, which was applied to map the top of an aquitard in a regional CPT dataset. Comparison with the more traditional time-consuming manually interpreted results suggest that this methodology is highly efficient, and should at least be used to support manual interpretation based on literature SBT classifications that provide indications on lithology, but lack information on the true typology of the data.

This methodology was only tested on a single horizon within the lithostratigraphical column. Further research should focus on the joint mapping of different horizons and layers. Moreover, to make the identification of the horizons more robust, constraints could be placed on the differences between identified depths at nearby CPTs, including a priori knowledge on the layer geometries.

References

Bhattacharya, B., Solomatine, D. P., 2006. Machine learning in soil classification. *Neural networks* **19**(2): 186-95. doi:10.1016/j.neunet.2006.01.005

Beerten, K., Wemaere, I., Gedeon, M., Labat, S., Rogiers, B., Mallants, D., Salah, S., Leterme, B. (Ed.), 2010. Geological , hydrogeological and hydrological data for the Dessel disposal site. Project near surface disposal of category A waste at Dessel. STB-SIE(HYD) - Version 1, NIROND-TR 2009-05 E (p. 261).

Beerten, K., Deforce, K., Mallants, D., 2012. Landscape evolution and changes in soil hydraulic properties at the decadal, centennial and millennial scale: A case study from the Campine area, northern Belgium. *Catena* **95**: 73–84. doi:10.1016/j.catena.2012.03.005

Bilski, P., Rabarijoely, S., 2009. Automated soil categorization using the CPT and DMT investigations. Int. Conf. on New Developments

in Soil Mechanics and Geotechnical Engineering (pp. 368-375).

Cetin, K. O., Ozan, C., 2009. CPT-Based Probabilistic Soil Characterization and Classification. *Journal of Geotechnical and Geoenvironmental Engineering* **135**(1): 84. doi:10.1061/(ASCE)1090-0241(2009)135:1(84)

Das, S., Basudhar, P., 2009. Utilization of self-organizing map and fuzzy clustering for site characterization using piezocone data. *Computers and Geotechnics* **36**(1-2): 241-248. Elsevier Ltd. doi:10.1016/j.compgeo.2008.02.005

De Craen, M., Beerten, K., Honty, M., Gedeon, M., 2012. *Geo-scientific evidence to support the I2 isolation function (geology & long-term evolution) as part of the Safety and Feasibility Case 1 (SFC1), SCK•CEN-ER-184* (p. 120). Mol, Belgium.

Dietrich, P., Leven, C., 2009. Direct push-technologies. In R. Kirsch (Ed.), Groundwater geophysics, a tool for hydrogeology (2nd ed., pp. 347-366). Springer. doi:10.1007/978-3-540-88405-7_12

DOV, 2011. Databank ondergrond Vlaanderen, 2011.

Facciorusso, J., Uzielli, M., 2004. Stratigraphic profiling by cluster analysis and fuzzy soil classification from mechanical cone penetration tests. In A. Viana da Fonseca & P. W. Mayne (Eds.), Proceedings ISC-2 on Geotechnical and Geophysical Site Characterization (pp. 905-912). Rotterdam: Millpress.

Fenton, G.A., 1999. Random field modeling of CPT data. *Journal of geotechnical and geoenvironmental engineering* **125**(6): 486-498.

Flach, G. P., Harris, M. K., Smits, A. D., Syms, F. H., 2005. Modeling aquifer heterogeneity using cone penetration testing data and stochastic upscaling methods. *Environmental Geosciences* **12**(1): 1-15. doi:10.1306/eg.06140404025

Forgey, E., 1965. "Cluster Analysis of Multivariate Data: Efficiency vs. Interpretability of Classification". In: *Bio- metrics* **21**: 768–769.

Fraley, C., Raftery, A. E., 2002. Model-Based Clustering, Discriminant Analysis, and Density Estimation. *Journal of the American Statistical Association* **97**(458): 611-631. doi: 10.1198/016214502760047131.

Fraley, C., Raftery, A. E., 2009. MCLUST Version 3 for R : Normal Mixture Modeling and Model-Based Clustering. (p. 56). Seattle. doi: Technical Report No. 504.

C. Fraley, A. E. Raftery, T. B. Murphy, L. Scrucca, 2012. mclust Version 4 for R: Normal Mixture Modeling for Model-Based Clustering, Classification, and Density Estimation, Technical Report No. 597, Department of Statistics, University of Washington, June 2012.

Gabriel, K. R., 1971. The biplot graphical display of matrices with applications to principal component analysis. *Biometrika* **58**: 453–467.

Goovaerts, P., 1997. Geostatistics for natural resources evaluation. Oxford University Press.

Gower, J.C., Hand, D. J., 1996. Biplots. Chapman & Hall.

Gulinck, M., 1962. Hydrogéologie II. Gisements aquifères liés aux formations Tertiaires et Quaternaires. Une carte au 1:500.000. *Atlas de Belgique*, Pl. 16B.

Gullentops, F., Wouters, L., (Eds.), 1996. Delfstoffen in Vlaanderen. Ministerie Vlaamse Gemeenschap, Brussel, 198pp.

Gullentops, F., Bogemans, F., De Moor, G., Paulissen, E., Pissart, A., 2001. Quaternary lithostratigraphic units (Belgium). *Geologica Belgica* **4**(1-2): 153–164.

Hartigan, J. A., Wong, M. A., 1979. A K-means clustering algorithm. *Applied Statistics* **28**: 100–108.

Hegazy, Y. A., Mayne, P. W., 2002. Objective Site Characterization Using Clustering of Piezocone Data. *Journal of Geotechnical and Geoenvironmental Engineering* **128**(12): 986-996. doi:10.1061/(ASCE)1090-0241(2002)128:12(986)

Ishioka, T., 2000. Extended K-means with an efficient estimation of the number of clusters. IDEAL '00: Proceedings of the Second International Conference on Intelligent Data Engineering and Automated Learning, Data Mining, Financial Engineering, and Intelligent Agents.

Ishioka, T., 2005. An expansion of x-means for automatically determining the optimal number of clusters — progressive iterations of k-means and merging of the clusters —. Proceedings of the Fourth IASTED International Conference, Computational Intelligence (pp. 91–96).

Jaksa, M. B., Kaggwa, W. S., Brooker, P. I., 1993. Geostatistical modelling of the spatial variation of the shear strength of a stiff , overconsolidated clay. In Li & Lo (Eds.), Probabilistic Methods in Geotechnical Engineering (pp. 185-194).

Jaksa, M., Brooker, P., Kaggwa, W., 1997. Modelling the spatial variability of the undrained shear strength of clay soils using geostatistics. In E. Baafi & N. Schofield (Eds.), Geostatistics Wollongong '96 (Vol. 2, pp. 1284-1295). Kluwer Publishers.

Jefferies, M.G., Davies, M.P., 1993. Use of CPTU to estimate equivalent SPT N60. *Geotechnical Testing Journal, ASTM*, **16**(4): 458-468.

Kulatilake, P., Um, J.-G., 2003. Spatial variation of cone tip resistance for the clay site at Texas A&M University. *Geotechnical and Geological Engineering*, **21**: 149-165.

Kurup, P. U., Griffin, E. P., 2006. Prediction of Soil Composition from CPT Data Using General Regression Neural Network. *Journal of*

Computing in Civil Engineering, **20**(4): 281. doi:10.1061/(ASCE)0887-3801(2006)20:4(281)

Kurup, P. U., Griffin, E. P., Tumay, M. T., 2010. Novel methodologies for soil characterization from CPT data. cpt'10.

Laga, P., Louwye, S., Geets, S., 2001. Paleogene and Neogene lithostratigraphic units (Belgium). *Geologica Belgica* **4**(1-2): 135–152.

Lenz, J. A., Baise, L. G., 2007. Spatial variability of liquefaction potential in regional mapping using CPT and SPT data. *Soil Dynamics and Earthquake Engineering* **27**: 690-702. doi:10.1016/j.soildyn.2006.11.005

Liu, C. N., Chen, C.-H., 2010. Spatial correlation structures of CPT data in a liquefaction site. *Engineering Geology* **111**(1-4): 43-50. Elsevier B.V. doi:10.1016/j.enggeo.2009.12.002

Lloyd, S., 1982. "Least Squares Quantization in PCM". In: IEEE Trans. Information Theory.

Louwye, S., De Schepper, S., Laga, P., Vandenberghe, N., 2007. The Upper Miocene of the southern North Sea Basin (northern Belgium): a palaeoenvironmental and stratigraphical reconstruction using dinoflagellate cysts. *Geological Magazine* **144**(1): 33.

Lunne, T., Robertson, P.K., Powell, J. J. M., 1997. Cone Penetration Testing in Geotechnical Practice. Blackie Academic and Professional, London, pp. 312.

MacQueen, J. B., 1967. "Some Methods for classification and Analysis of Multivariate Observations, Proceedings of 5-th Berkeley Symposium on Mathematical Statistics and Probability", Berkeley, University of California Press, 1:281-297.

Młynarek, Z., Wierzbicki, J., Wołyński, W., Tschuschke, W., 2008. Assessment of Efficiency of Different Cluster Analysis Methods for Evaluation of a Stratigraphy of Strongly Laminated Subsoil. The 12th International Conference of International Association for Computer Methods and Advances en Geomechanics (IACMAG) (pp. 1291-1299). Goa, India.

ONDRAF/NIRAS. 2010. *Het cAt-project in Dessel. Een langetermijnoplossing voor het Belgische* categorie A-afval. Retrieved from http://www.niras-cat.be/downloads/cAt_masterplan_NL_LOW.pdf on 07-12-2011.

Pebesma, E.J., 2004. Multivariable geostatistics in S: the gstat package. *Computers & Geosciences* **30**: 683-691.

Pelleg, D., Moore, A., 2000. X-means: Extending k-means with efficient estimation of the number of clusters. Proceedings of the Seventeenth International Conference on Machine Learning.

Robertson, P., Campanella, R., Gillespie, D., Greig, J., 1986. Use of piezometer cone data. IN-SITU '86 ASCE Specialty Conference on Use of In-situ Testing in Geotechnical Engineering.

Robertson, P. K., 1990. Soil classification using the cone penetration test. *Canadian Geotechnical Journal* **27**: 151-158.

Robertson, P. K., 2010. Soil behaviour type from the CPT : an update. cpt'10.

Robertson, P., Cabal, K., 2010. Guide to Cone Penetration Testing for Geotechnical Engineering (4th ed., p. 138). Gregg Drilling & Testing, Inc.

Robertson, P.K., Wride, C.E., 1998. Evaluating cyclic liquefaction potential using the cone penetration test, *Canadian Geotechnical Journal* **35**: 442 – 459.

Rogiers, B., Schiltz, M., Beerten, K., Gedeon, M., Mallants, D., Batelaan, O., Dassargues, A., Huysmans, M., 2010. Groundwater model parameter identification using a combination of cone-penetration tests and borehole data. IAHR international groundwater symposium (p. 19). Valencia.

Rogiers, B., Beerten, K., Smeekens, T., Mallants, D., Gedeon, M., Huysmans, M., Batelaan, O., Dassargues, A., 2013a. The usefulness of outcrop analogue air permeameter measurements for analysing aquifer heterogeneity: Quantifying outcrop hydraulic conductivity and its spatial variability. *Hydrological processes*. DOI: 10.1002/hyp.10007

Rogiers, B., Beerten, K., Smeekens, T., Mallants, D., Gedeon, M., Huysmans, M., Batelaan, O., Dassargues, A., 2013b. Derivation of flow and transport parameters from outcropping sediments of the Neogene aquifer, Belgium. *Geologica Belgica*, **16**(3): 129–147.

Schiltz, M., 2008. Lithological and Stratigraphical interpretation by means of cone penetration tests (CPT's) in the Dessel-Kasterlee-Geel-Mol area. Bvba SAMSUFFIT Geoservices, Fieldsurvey cAt 2008.

Schiltz, M., 2010. Lithological and Stratigraphical interpretation of cone penetration tests (CPT's) executed for the first tumulus at the disposal site in Dessel and in the Dessel-Kasterlee-Geel-Mol area. Bvba SAMSUFFIT Geoservices, Fieldsurvey cAt 2010.

Sibelco, 2010. Silica sand of Mol. Technical Datasheet TDS.03.05.10 2010-12-09

Telgarski, M., Vattani, A., 2010. Hartigan's Method: k-means Clustering without Voronoi. Proc. of 13th International Conference on Artificial Intelligence and Statistics (AISTATS) (Vol. 9, pp. 820-827). Chia Laguna Resort, Sardinia, Italy.

Tillmann, A., Englert, A., Nyari, Z., Fejes, I., Vanderborght, J., Vereecken, H., 2008. Characterization of subsoil heterogeneity, estimation of grain size distribution and hydraulic conductivity at the Krauthausen test site using Cone Penetration Test. *Journal of contaminant*

hydrology **95**(1-2): 57-75. doi:10.1016/j.jconhyd.2007.07.013

Uzielli, M., Vannucchi, G., Phoon, K., 2005. Random field characterisation of stress-normalised cone penetration testing parameters. *Geotechnique* **55**(1): 3–20.

Vermeiden, J., 1948. Improved sounding apparatus as developed in Holland since 1936. Proc. 2nd Int. Conf. on Soil Mech. and Found. Eng., Rotterdam, 1: 280-287.

Wouters, L., Schiltz, M., 2012. Overview of the field investigations in and around the nuclear site of Mol-Dessel, NIROND-TR 2011-42.

Zhang, Z., Tumay, M. T., 1999. Statistical to fuzzy approach toward CPT soil classification. *Journal of Geotechnical and Geoenvironmental Engineering* **125**(3): 577-580.

Chapter 8

Multi-scale aquifer characterization and groundwater flow model parameterization using direct push technologies

Based on Rogiers B, Vienken T, Gedeon M, Batelaan O, Mallants D, Huysmans M, Dassargues A. Multi-scale aquifer characterization and groundwater flow model parameterization using direct push technologies. Environmental Earth Sciences, submitted.

Abstract

Direct push technologies (DPT) are typically used for cost-effective geotechnical characterization of unconsolidated soils and sediments, but recent developments also allow for hydraulic conductivity characterization. Up to now, however, only a limited number of studies document such continuous *in situ* data to develop three-dimensional conceptual models and parameterize groundwater flow models. This study demonstrates how DPT improve conceptual model building and evaluates the degree to which DPT-derived hydrogeologic parameters measured at different spatial scales improve performance of a regional groundwater flow model. The study area covers an area of ~60 km² including the nuclear zone of Mol/Dessel, Belgium. The hydrostratigraphy was obtained from analysis of cored boreholes and several hundred cone penetration tests (CPTs). Hydrogeological parameters were derived from the same cores and CPTs and also from hydraulic direct push tests. Multiple realizations of the three-dimensional subsurface were generated using a non-stationary multivariate geostatistical approach whereby the groundwater model parameter realizations were conditioned on the borehole and direct push data. After the selection of the best out of 20 random realizations (based on reproduction of heads and head differences), and local global parameter optimization, a significant increase in model performance (in terms of head observations and vertical head differences) was obtained when compared to a reference case with homogeneous geological layers. Comparison of solute transport simulations based either on the optimal or reference groundwater model flow fields demonstrate the impact that accounting for subsurface heterogeneity through DPT can have on the quantification of solute transport. We demonstrate that by using DPT, conceptual model uncertainty decreases and flow model performance increases due mainly to improved estimation of *K* variability from calibrated CPT data. The use of hydraulic direct push tools is more time consuming, but a higher accuracy can be obtained with similar data resolution, and resulted in this case in the detection of a high *K* layer that was not considered before in the hydrogeological conceptual model.

8.1. Introduction

Groundwater flow and contaminant transport models are commonly used to support decision making regarding waste disposal options (Patyn *et al.*, 1989; Gedeon *et al.*, 2011; Gedeon & Mallants 2012; Selroos & Painter 2012), sites contaminated by surface or subsurface sources (*e.g.* Hammond & Lichtner 2010; Huysmans *et al.*, 2006, Battle-Aguilar *et al.*, 2009), or to develop and test cost-effective groundwater remediation or monitoring schemes (*e.g.* Saito & Goovaerts 2003; Bierkens 2005). Such models are influenced by different sources of uncertainty (Nilsson *et al.*, 2007; Christian *et al.*, 2006, 2007; Rojas *et al.*, 2008), including those due to spatial variability in aquifer and aquitard properties as hydraulic conductivity (*K*; a list of symbols and abbreviations is provided in Table 7.1). Quantifying spatial variability of

K remains challenging. Classical drilling techniques for shallow heterogeneous unconsolidated sedimentary deposits involving continuous coring are expensive and time-consuming, especially when the area of interest exceeds several tens of km².

Alternative techniques such as direct push technologies use hydraulic rams, supplemented with vehicle weight, or high-frequency hammering, to advance small-diameter tools into the subsurface (*e.g.* Dietrich & Leven 2006; McCall *et al.*, 2006). These tools are typically used for cost-effective geotechnical characterization of unconsolidated deposits using cone penetration tests (CPTs; Lunne *et al.*, 1997), with typically a vertical measurement spacing of ~2 cm. In hydrogeological studies, such cone penetration testing data is mainly used for indirect lithostratigraphic profiling, but some studies report the use of this data to obtain hydraulic conductivity estimates. This is mainly achieved by using soil behavior types (SBTs; *e.g.* Robertson 2010; Rogiers *et al.*, 2010) or indirectly through other variables (*e.g.* Flach *et al.*, 2005; Tillmann *et al.*, 2008), yielding a set of discrete *K* estimates. Predicting continuous *K* estimates has been done for piezocone sounding (CPTu; *e.g.* Elsworth & Lee 2007; Lee *et al.*, 2008; Chai *et al.*, 2011) and the corresponding dissipation testing (*e.g.* Teh & Houlsby 1991; Van Baars & Van De Graaf 2007; Robertson 2010; Chai *et al.*, 2012) but is seldom performed for standard CPT data.

More recent but less established developments like direct push injection logging (DPIL; Dietrich *et al.*, 2008; Liu *et al.*, 2009; Lessof *et al.*, 2010) and the hydraulic profiling tool (HPT; GEOPROBE 2007; McCall *et al.*, 2009; Köber *et al.*, 2009) also allow for high-resolution hydraulic characterization with direct push equipment, by which relative *K* values are obtained. The more standard direct push slug testing (DPST; Butler 2002; Butler *et al.*, 2002) or the direct push permeameter (Butler *et al.*, 2007) are often used for calibration of the obtained high-resolution relative *K* data (*e.g.* Dietrich *et al.*, 2008; Liu *et al.*, 2009; Lessof *et al.*, 2010).

The depth of investigation of these direct push tools depends on the equipment used (*i.e.* applied load) and sediment properties (friction). In sandy unconsolidated sediments, depths of ~40 m can be reached. Up to now however, most of these methods are applied at small test sites, and only a limited number of studies document using this type of data on a more regional scale, *e.g.* to parameterize regional groundwater flow models. To the best of our knowledge, the study by Flach *et al.* (2005) represents the most large-scale application (8 km²) of direct push data for groundwater flow model parameterization. These authors succesfully used CPTu data for deriving four discrete *K* classes based on fines content, estimated from the CPTu data, and normalized pore pressure. A threshold for the latter parameter was used to define an extra low *K* class. Given the absence of pore pressure data with standard CPT tests, the parameterization of an aquitard using this type of data is particularly challenging. Moreover, as adoption of the hydraulic direct push tools by the environmental industry only grows slowly, demonstration of the usefulness and effectiveness of such data, and compatibility with geotechnical and borehole data is highly required.

To fill these gaps, this study aims at parameterising a regional groundwater flow model using data from various types of direct push technologies, and investigating the effect on the model performance. To achieve this, the usefulness of CPTs for continuous hydraulic conductivity estimation is explored, and a geostatistical methodology is used to integrate this secondary data with the primary calibrated hydraulic direct push and borehole data. The approach is applied to an area of ~60 km² near the nuclear zone of Mol/Dessel, Belgium, and part of the Neogene aquifer, where different direct push datasets are available, as well as seven cored boreholes down to a depth of 50 m (Figure 8.1A-D), all performed in the framework of the ONDRAF/NIRAS (2011) cAt project for the disposal of low- and intermediate-level short-lived (category A) waste. The Mio-Pliocene geological formations characterized by the different direct push campaigns are the Mol, Kasterlee and Diest Formations, which are covered by Quaternary sediments of a few meter thickness. The main lithologies of these formations include fine to medium grained, glauconitic, micaceous sands with Fe sandstone layers and lignite layers. A varying clay content is found in certain units (mainly Kasterlee and Diest Formations), while basal gravels are present between the units (Laga *et al.*, 2001). Deposition took place in a shallow marine to perimarine environment at the southern margin of the North Sea Basin (Louwye *et al.*, 2007; Louwye & Laga, 2008;

Louwye & De Schepper, 2010). A lithostratigraphical profile through the Neogene aquifer in the study area is shown in Figure 8.2. Hydrogeologically, the Neogene aquifer is divided in an upper aquifer containing the Quaternary, the Mol Upper, Mol Lower and Kasterlee Sands, an aquitard corresponding to the Kasterlee Clay, and a lower aquifer containing the clayey top of the Diest Formation, and the Diest, Dessel, Berchem and Voort Sands. The lower boundary of the Neogene aquifer is represented by the top of the Rupelian Boom Clay aquitard (Yu *et al.*, 2013).
This work is a follow-up of the study by Rogiers *et al.* (2012b), who used only the CPT data within the Kasterlee Clay aquitard in a relative way to optimize the aquitard *K* field, with a resulting decrease of the sum of squared errors to 35% of the reference value. Given the positive outcome, the approach is now extended to the entire upper ~40 m of the subsurface with absolute CPT-based *K* estimates, and newly available borehole core data (Rogiers *et al.*, 2012c, 2013b; Chapter 5) and hydraulic direct push data (Vienken *et al.*, 2012; Tinter 2012) targeting the uppermost litho-stratigraphical layers.

8.2. Methods

In this study two sets of independent data are combined to inform conceptual model building and to constrain a numerical groundwater flow model. At the scale of the entire model domain (~60 km²), approximately 265 standard CPT logs were used for derivation of continuous hydraulic conductivity depth profiles, taking detailed laboratory and air permeability measurements on seven cored boreholes as calibration data (Rogiers *et al.*, 2013b; Chapter 5). At the scale of the future disposal site (~0.02 km^2), a small number of hydraulic direct push tests were performed including both relative (Direct Push Injection Logging, Hydraulic Profiling Tool) and absolute (Direct Push Slug Tests) *K* information. The absolute measurements allow calibration of the relative data, thus producing additional high-resolution vertical *K* profiles. We first discuss the three hydraulic direct push testing methodologies and their interrelationships; next we demonstrate the conversion of the geotechnical direct push tests into continuous *K* profiles using borehole core measurements for calibration.

8.2.1. Hydraulic direct push tests

The hydraulic direct push testing was targeting the uppermost lithostratigraphical units of the future site for surface disposal of low-level radioactive waste (Figure 8.1C; ONDRAF/NIRAS, 2010). High spatial resolution *K* profiles are needed at this site to develop reliable models of groundwater flow and solute transport as part of the performance assessment of the future disposal facility. In total 17 Direct Push Injection Loggings (DPIL), 6 Hydraulic Profiling Tool (HPT) tests and 19 Direct Push Slug Tests (DPST) were performed along two quasi orthogonal transects (Figure 8.1D). Examples of the data obtained at five locations up to depths of approximately 15 m are displayed in Figure 8.3. The relative *K* values (K_{rel}) obtained by DPIL and HPT will further be transformed into absolute *K* values using the DPST (K_{DPST}) data for calibration.
The hydraulic direct push tests clearly image a high *K* zone (Figure 8.3), which corresponds to the coarse white Upper Mol Sands, most typical lithology for the continental/estuarine Mol Formation, in the Dessel-2 borehole (Beerten *et al.*, 2010), where only few measurements on this lithology were performed. This high *K* layer was not accounted for before, and likely occurs at a regional scale, east of our study area.

8.2.2. Geotechnical direct push tests

CPT data was gathered in two campaigns: in 2008 a total of 180 CPT tests were carried out in a study area of ~60 km² based on a grid with spacing of ~500-800 m (see Figure 8.1B). In 2010 a smaller area of ~200 x 400 m² was involved in collecting 85 CPT profiles (see Figure 8.1C); seven cored boreholes with depths between 40 and 50 m were also retrieved. At these seven boreholes co-located CPT data were used to calibrate the latter for estimating *K*. CPT depth profiles ranged between 15 and 42 m, with 60% of the soundings over 30 m. Most of the 265 CPTs were standard CPT tests, without registration of pore pressure. The influence of correcting for pore pressure using a standard appraoch (Lunne *et al.*, 1997) was tested for about 40 CPTu logs, but was proven to be negligible (results not shown). The cone area for all CPT tests was 1500 mm².

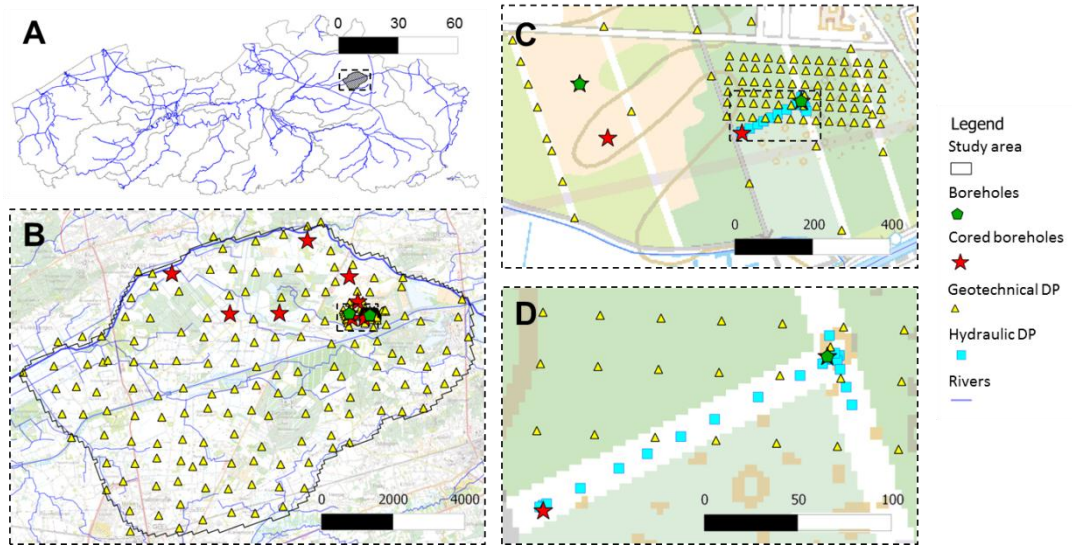

Figure 8.1: A) Situation of the study area within the hydrography of Flanders, Belgium. B) Overview of drilling and DPT sites within the study area. C) drilling and DPT sites at the future low-level waste disposal site (ONDRAF/NIRAS, 2010). D) location of the hydraulic direct push tests.

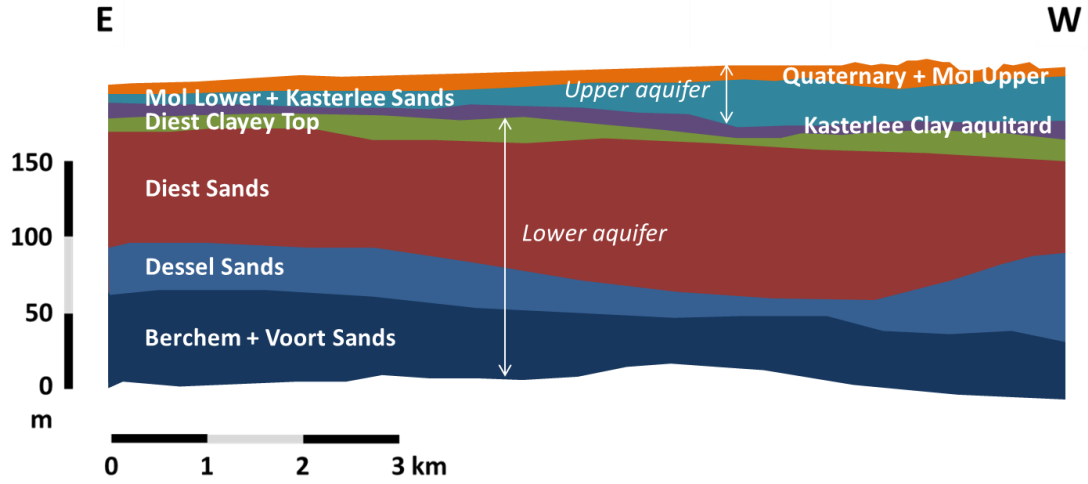

Figure 8.2: East-west lithostratigraphical profile through the study area.

Table 8.1: List of symbols and abbreviations.

Symbol	Explanation
K	Hydraulic conductivity
CPT	Cone Penetration Test
CPTu	Piezocone test
DPIL	Direct Push Injection Logging
HPT	Hydraulic Profiling Tool
DPST	Direct Push Slug Test
SBT	Soil behaviour type
Q_{tn}	Normalized cone resistance (MPa)
F_r	Normalized friction ratio (%)
z_{strat}	Stratigraphical depth
z_{masl}	Depth in meter above sea level
BE_{Fr}	Friction ratio boundary energy
I_c	Soil behaviour type index
PC	Principal Component
RMSE	Root mean squared error

In a previous study by Rogiers *et al.* (2012b) the aquitard (Kasterlee Clay aquitard) separating the main aquifer units within the Neogene aquifer had its *K* field parameterized in a relative way by using a standardized CPT-based heterogeneous parameter field and local optimization of the mean and variance of the aquitard. This was done because of the inability to capture by means of CPT the discrete clay lenses exhibiting very low *K* values, thus causing a bias in the aquitard *K* estimates. This is a consequence of the too large support volume of an individual CPT measurement (the measurement can be influenced by sediments up to several cone diameters distance from the CPT cone) compared to the limited thickness of the clay lenses (a few centimetres at most). Small excursions of the signal are however visible. To overcome this limitation in achieving a high spatial granularity of the measurement, Bhattacharya & Solomatine (2005) proposed to use a shape characteristic of the CPT log, the so-called boundary energy, for translating the log characteristics into a soil classification using supervised (by expert interpretations) machine learning. Boundary energy is defined as the amount of energy required to modify the shape of a contour (*i.e.* the log of a CPT parameter) to its lowest energy level, which is a circle, with the same perimeter as the original object. While these authors used a multi-scale representation of boundary energy, we used a simplified finite difference approximation of the curvature and calculated the friction ratio boundary energy as

$$\mathrm{BE}_{F_r} = \frac{1}{N}\sum_{n=1}^{N}\left(\frac{\left|\frac{d^2F_{r,n}}{dz^2}\right|}{\left(1+\left(\frac{dF_{r,n}}{dz}\right)^2\right)^{3/2}}\right)^2 \qquad [8.1]$$

with N points within a certain range z of the point where BE_{Fr} is calculated, and $F_{\mathrm{r,n}}$ the friction ratio at point n. For the maximum z range we used 0.10 m, which corresponds to averaging the squared curvature over 11 points with a CPT data spacing of 0.02 m. This should result in a clear signal for clay lenses of 5 to 10 cm thick. The resulting BE_{Fr} values are displayed in Figure 8.4C for the Dessel-2 borehole, which clearly shows how such sudden but small excursions of the standard CPT logs (Figure 8.4B) are captured by the BE_{Fr} signal.

Rogiers *et al.* (2012b) tested the usefulness of the BE_{Fr} parameter to capture the Kasterlee Clay aquitard clay lenses by using a cutoff value for the BE_{Fr} signal to identify the two lithological classes sand and clay. After assigning representative *K* values to each of the two classes followed by vertical upscaling, the resulting groundwater flow model did only minimally improve its performance compared to a model with uniform aquitard *K* (relative SSE decreased by 4%).

Rogiers *et al.* (2013b; Chapter 5) further derived high-resolution *K* logs (*e.g.* Figure 8.4A) using air permeability measurements on the same borehole cores used in this study. This allows a more data-driven approach to investigate the use of the boundary energy for capturing the effect of the clay lenses more carefully. Visual comparison of the high-resolution borehole log with the BE_{Fr} parameter in Figure 8.4 already indicates this parameter is a good indicator for thin clay lenses.

8.2.3. High resolution data calibration

Calibration of respectively the geotechnical direct push and the hydraulic direct push test data occurs in different independent ways. For calibrating the geotechnical CPT data high-resolution *K* logs from borehole cores are used; the hydraulic direct push data calibration is by means of Direct Push Slug Tests.

Different approaches exist to calibrate the relative conductivities K_{rel} obtained from DPIL or HPT with the absolute conductivities (K_{DPST}) measured by the Direct Push Slug Tests. Assuming the measurement support discrepancies between the absolute (*e.g.* due to the DPST screen length) and the relative *K* measurements can be neglected, the K_{rel} values are rescaled using a power law with standard regression techniques, as applied by Dietrich *et al.* (2008):

$$K_{DPST} = a \times K_{rel}^{b} \qquad [8.2]$$

where *a* and *b* are calibration parameters. Liu *et al.* (2009) extended this approach by using a numerical model for simulating Direct Push permeameter test responses based on the rescaled relative *K* data, and used inverse modelling to obtain the optimal *a* and *b* parameters. When DPST screen lengths become much longer than the lengths over which measurements occur for DPIL and HPT, measurement support discrepancies cannot be neglected. In such case, upscaling of the relative *K* data can be performed prior to calibration if

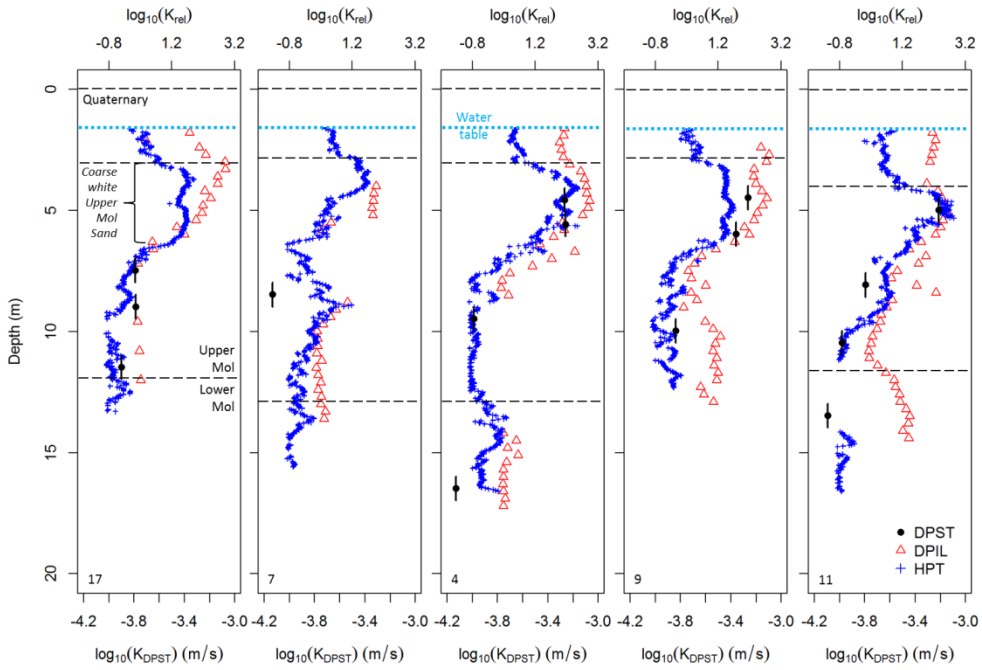

Figure 8.3: Examples of hydraulic direct push logs with approximate locations of the water table and the main lithostratigraphical boundaries based on the Dessel-2 borehole interpretations (Beerten *et al.*, 2010). K_{rel} = relative K from Direct Push Injection Logging (DPIL; l/h/bar) or Hydraulic Profiling Tool (HPT; ml/min/Pa); K_{DPST} = absolute K (m/s) from Direct Push Slug Test (DPST).

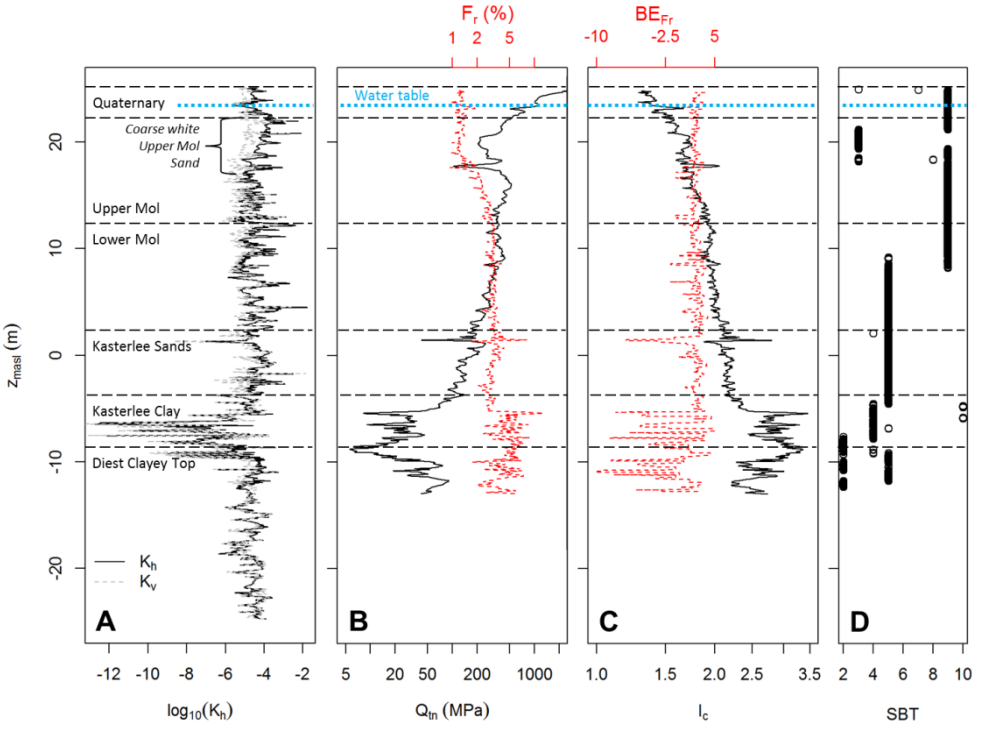

Figure 8.4: Example of A) high-resolution borehole core *K* logs (Dessel-2 borehole; Rogiers *et al.*, 2013b; Chapter 5), B) the corresponding standard normalized CPT data, C) CPT- derived soil behaviour type index I_c and the friction ratio boundary energy BE_{Fr} and D) site-specific soil behaviour types (SBT) obtained using the model-based clustering approach of Rogiers *et al.* (2013c; Chapter 7) with I_c and z_{strat} as variables.

the relative K data spacing is small compared to the screen length used for the absolute K measurements, as proposed by Lessof *et al.* (2010). When vertical integral scales are only a fraction of the horizontal integral scales, arithmetic and harmonic means can be used for upscaling. If the number of relative K measurements within the length of the screen used for measuring calibration data is small, a better calibration outcome is pursued here by using ordinary kriging and sequential gaussian simulation along the screen prior to upscaling. This is especially useful for the DPIL logs, where a typical data spacing of 0.3 m is used.

We tested the following upscaling approaches for DPIL and HST measurements: 1) applying Eq. 8.2 with parameters a and b without upscaling, 2) upscaling with ordinary kriging or sequential gaussian simulation for estimating K at each cm along the DPST filter length (1.06 m in our case), and 3) taking the arithmetic mean for obtaining a horizontal K value (K_h), which can be compared to the corresponding horizontal K_{DPST} values. The mismatch between calculated and measured K_{DPST} values was minimized iteratively using the simplex algorithm of Nelder & Mead (1965). The use of sequential Gaussian simulation instead of interpolation by kriging for obtaining a DPST screen support estimate was tested but the gain proved to be negligible and was not further applied (results not shown). By using the power law prior to upscaling, we enable the use of harmonic means to obtain vertical K values (K_v) as well, at any given vertical scale. The obtained anisotropy (K_h/K_v) is then an additional source of information for conditioning the groundwater flow model.

In calibrating the CPT data, the use of standard CPT data (Figure 8.4B) was tested first. However, the Q_{tn} signal (normalised cone resistance) is negatively correlated with the borehole K logs for the coarse white Upper Mol Sand in Figure 8.4A, while systematic postivite correlation is present for the remainder of the lithostratigraphical column. Furthermore, a non-linear data-driven modelling approach by Rogiers *et al.* (2012a; Chapter 6) was not able to correct for this as the high K Upper Mol Sand only occurs at the very top of one of the boreholes. Therefore, we tested the soil behaviour type index I_c by Robertson & Wride (1998)

$$I_c = ((3.47 - log\ Q_{tn})^2 + (log\ F_r + 1.22)^2)^{0.5} \quad [8.3]$$

with Q_{tn} and F_r the normalized cone resistance and friction ratio, respectively. I_c is essentially the radius of concentric circles that can be plotted in the soil behavior type diagram by Robertson (1990). A power law as in Eq. 8.2 was then applied to the I_c data obtained with Eq. 8.3 while adding an additional correction using BE_{Fr}:

$$K_{CPT} = exp\left(ln\left(c \times I_c^d\right) - e_{SBT} \times \left(ln\left(BE_{F_r}\right) - f_{SBT}\right)\right) \quad [8.4]$$

with K_{CPT} the estimated hydraulic conductivity, c and d rescaling parameters for I_c, and e_{SBT} and f_{SBT} are additional parameters which depend on the soil behaviour type (SBT) obtained through site-specific SBT classification of the CPT data using the $I_c - z_{strat}$ approach of Rogiers *et al.* (2013c; Chapter 7; Figure 8.4D). Including dependence on the SBTs is required as the CPT signal within the lower aquifer gives rise to BE_{Fr} peaks without the presence of clear discrete clay lenses. This could be caused by the presence of concretions as the Diest Formation is known for containing iron sandstones. After the corrections introduced via Eq. 8.4, both the CPT-based K estimates (K_{CPT}) and the high-resolution borehole K logs are upscaled for each section of one meter long with arithmetic and harmonic means, to avoid strong influences of small errors in the vertical position of both logs, and a final linear model was derived for upscaled CPT-based K_h and K_v values using upscaled borehole K_h and K_v for calibration. The parameters a, b, c_{SBT}, d_{SBT} and upscaled linear model parameters are all iteratively optimized using the Nelder & Mead (1965) approach. To test no overfitting occurred in optimising parameters for K_{DPST} and K_{CPT}, a leave-one-out (one DPST or one meter of borehole core log data) cross-validation is performed for all data calibration cases.

8.2.4. Development and conditioning of the groundwater flow model

The groundwater flow model used in this paper is an adaptation of the steady-state MODFLOW-2005 (Harbaugh 2005) model developed by Gedeon *et al.* (2011). Sensitivity analysis for this model was performed by Gedeon & Mallants (2012). Rogiers *et al.*

(2012b) used it for testing different concepts of parameterization of the Kasterlee Clay aquitard based on the CPT dataset without any further model modifications. The original model will be modified here to evaluate the degree and nature of improvement in model performance when different additional model parameter sets are used.

Model boundary conditions in the upper model layer are unchanged in so far as river, drain, fixed and general head boundary conditions are concerned (Figure 8.5A). Groundwater recharge is modified from a spatially uniform value to a spatially variable recharge based on separate vertical flux modelling. For this purpose Leterme & Mallants (2012) applied the HYDRUS-1D simulator (Šimůnek 2005) and determined recharge for crops, meadow, deciduous and coniferous land cover types (Figure 8.5B). No recharge value for the built-up area was previously defined; here we calibrate it as a percentage of the meadow recharge. Other boundary conditions within the upper aquifer concern all layers below the upper layer; nearly the entire model area is enclosed by rivers for which no flow conditions are assumed. In the second aquifer below the aquitard, no flow conditions exist under the main river (Kleine Nete) which is assumed to drain the Neogene aquifer completely. At other locations, fixed head conditions are used based on head values obtained from the regional scale Neogene Aquifer Model (Gedeon 2008) (Figure 8.5C).

The 86 piezometers included in the Gedeon *et al.* (2011) model are now supplemented by 44 additional piezometers with filter depths of maximum 21 m below surface (Figure 8.5D shows all 130 piezometers). Average head values for the 86 piezometers were calculated from monthly observations between 1990 and 2010; the same was done for the 44 additional piezometers. Moreover, all piezometers are now implemented in MODFLOW as multi-layer observations, with weighted contributions of different cells to the mean piezometer head, where weights are defined by the Gedeon *et al.* (2011) model K_h values. This is necessary to include effects of small-scale heterogeneity on piezometer heads especially for a model with a large number of thin numerical layers. Applying this new approach with the reference parameterization resulted in a higher accuracy (*i.e.* smaller model error) for several of the piezometers. Two types of data were used for model calibration: 1) 130 average head values for piezometers in upper and lower aquifers, and 2) 68 vertical head differences obtained from different piezometers at the same location. The root mean squared error (RMSE) is used as objective function for all head measurements and head differences, and no weighting factors were applied for these two data types in the objective function.

The MODFLOW Hydrogeologic Unit Flow (HUF) package (Anderman & Hill 2000) was used for parameterization of the K_h and K_v data, independent of the numerical grid. The upper 38 m of the subsurface is treated differently from the remaining part of the model, as it includes most of the data for forward conditioning (Figure 8.6A), and is denoted as the HUF3D section (Figure 8.6B). The bottom of this section is obtained by smoothing the topography and substracting 38 m. In the north-eastern part of the study area, a few additional meters are included to encompass the entire Kasterlee Clay unit. The HUF layers within this section are all parallel to the top of the Kasterlee Clay, which is the most pronounced lithostratigraphical boundary in the system. This horizon was mapped manually by Schiltz (2008, 2010; see also Wouters & Schiltz 2012), and validated by Rogiers *et al.* (2013c; Chapter 7) who showed that it is traceable in an automatic objective way using site-specific soil behaviour type classification. Due to the inclination of the geological layers, 39 HUF layers of one meter thickness are needed above the Kasterlee Clay aquitard top, and 50 one-meter layers below it, to fill the entire HUF3D section. Where the layers extend beyond the HUF3D section (above ground surface or below the bottom of the HUF3D section), their thickness is set to zero. Parameterization of this section is conditioned on all direct push and borehole data, in 3D.

The remaining part of the model contains part of the Diest, the Dessel, Berchem and Voort/Eigenbilzen Formations, and is parameterized with 75 HUF layers of varying thickness based on the data density at the GT01, GT02 and D5 boreholes (Figure 8.6A), parallel to the layer bedding (Figure 8.6B). To parameterize this section, denoted as HUF1D, each layer has a constant K_h and K_v value derived from the available measurements from the Dessel-5 borehole (Beerten *et al.*, 2010; Labat *et al.*, 2011),

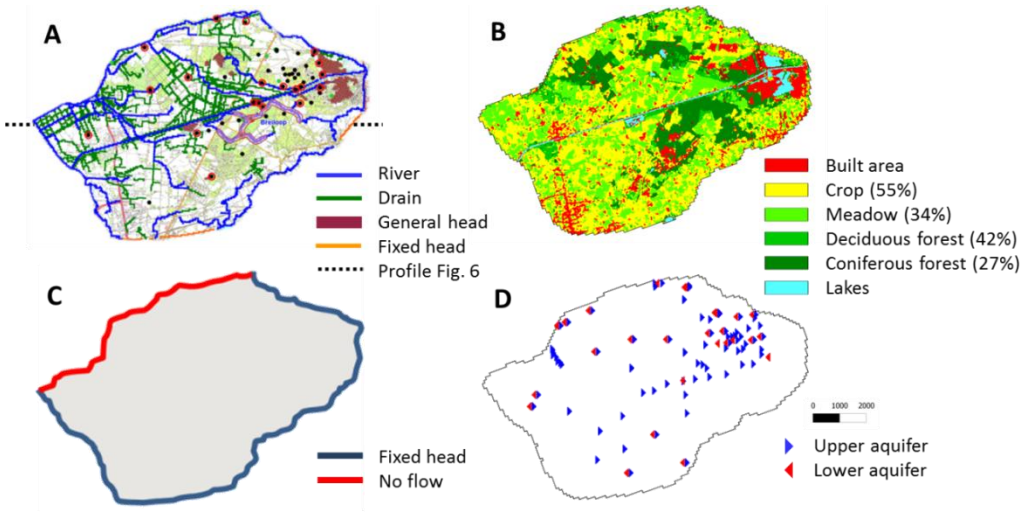

Figure 8.5: Boundary conditions (A-C) of the groundwater flow model, and location of the piezometers (D). A) represents the upper aquifer boundary conditions, B) the upper layer recharge in percentage of the total annual precipitation (899 mm; Leterme & Mallants 2012), and C) the lower aquifer boundary conditions (applicable to all model layers). The built-up area recharge is taken as a parameter for optimization, and the lakes correspond to the general head conditions in A).

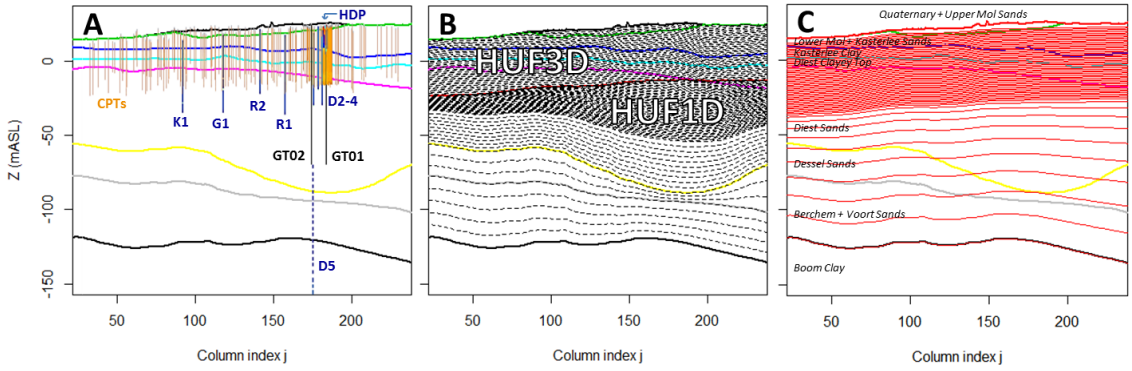

Figure 8.6: A) Litho-stratigraphy and projected borehole and direct push logs (HDP); D2-4, D5: Dessel-2, -3, -4 and -5 boreholes; G1: Geel-1 borehole; K1: Kasterlee-1 borehole; R1, R2: Retie-1 and -2 boreholes; GT01, GT02: geotechnical boreholes; HPD: hydraulic direct push tests; CPTs: ~265 cone penetration tests. B) Hydrogeological unit flow (HUF) package layers (dashed lines), used for the parameterization of K. HUF3D indicates the section in which parameterization is done in 3D; HUF1D indicates all layers have a constant value and parameterization for is by layer (1D vertical direction). C) Vertical numerical model discretization. The parameters provided for the HUF layers are mapped to the numerical layers using arithmetic and harmonic means.

and grain-size-based *K* predictions for the GT01 and GT02 boreholes, using the results obtained by the GLUE-ANN approach of Rogiers *et al.* (2012a; Chapter 6). Hence, data conditioning within this HUF1D section is done along the stratigraphical depth, in 1D.

Finally, the vertical numerical discretization of the model is adapted from 14 layers of 50x50 m² cells to 57 layers (Figure 8.6C). The upper layer thickness is mostly between 3 and 7.5 meter in order to assure presence of the water table within this layer. This avoids numerical problems with dry cells and convergence of the model. The following 46 layers all have a thickness of 1 m, and cover the entire HUF3D section, as well as the uppermost part of the HUF1D section in certain areas. The remaining 10 layers are used to fill up the remaining part of the model, with a linear increase in layer thickness. Translation of the HUF parameters to

the numerical grid is done automatically by MODFLOW. This enables us to work bedding-parallel for the parameterization of the model, while keeping numerical stability.
To condition hydraulic conductivity within the HUF3D section on the direct push and borehole data, the following assumptions were made:

1) All hydraulically obtained data (borehole *K* logs and hydraulic direct push tests) is considered as the primary variable, and the geotechnical direct push data is used as secondary variable.
2) As the amount of primary data is considerably lower than that of the secondary data, we assume their spatial statistics to be similar, and use the variograms obtained from the latter for both datasets. The cross-variogram is obtained by multiplying the direct variogram with the correlation between primary and secondary data, as calculated at the cored borehole locations.
3) As the studied Neogene aquifer domain is clearly non-stationary, a classical geostatistical approach cannot be used. A full 3D non-stationary approach is however computationally not feasible with the current implementation and processing capability. Therefore, we consider only non-stationarity in the vertical direction. Non-stationarity of the mean, the correlation coefficient between primary and secondary data, and the secondary data variograms is obtained by using distance weighting (Harris *et al.*, 2010; Machuca-Mory & Deutsch 2012). The data weights ω for this distance weighting at an anchor point *o* with *n* samples at locations u_α are obtained from the Gaussian kernel

$$\omega(u_\alpha; o) = \frac{\varepsilon + exp\left(-\frac{(d(u_\alpha;o))^2}{2s^2}\right)}{n\varepsilon + \sum_{\alpha=1}^{n} exp\left(-\frac{(d(u_\alpha;o))^2}{2s^2}\right)} \qquad [8.5]$$

with bandwidth *s*, background constant *ε*, to avoid computational problems, and $d(u_\alpha;o)$ the vertical distance in our case. After testing different bandwidths, we decided to use a value of 8 m, as this resulted in the most diversity between the different locations within the lithostratigraphical column. For the background constant we used a value of 0.01. These weights are used for calculating local mean, correlation and direct variograms along the lithostratigraphical column, and these are considered stationary each time within the bedding-parallel HUF layer.
4) To keep this multivariate non-stationary geostatistical simulation computationally feasible, we selected a subset of all target data locations (HUF cells) by only retaining each 5th row and column number (all layers are kept), which reduces the simulated data locations from 1,021,899 to 41,183. Inverse distance weighting with a power of 0.1 is used to extrapolate this data to the entire HUF3D section. Given the local stationarity assumption for all bedding-parallel HUF layers, a single layer can be simulated using standard sequential Gaussian simulation. Therefore, to speed up the simulation, all layers were simulated in a random order instead of using a random 3D path through all points. This simplification might induce slightly larger spatial correlation in the vertical direction, according to the observations of McLennan (2002). Finally, the 2010 CPT data is thinned to maximum one CPT test per cell at the disposal site. The secondary data at the borehole locations is removed as it is redundant due to the presence of primary data. To limit the amount of correlated variables in the geostatistical simulation, the logarithmic K_h and K_v data was orthogonalized using principal component analyses, and simulations were done independently for the principal components (PCs). For creating one realization of the HUF3D K_h and K_v parameters, the implementation of these assumptions in R, calling the gstat package (Pebesma 2004) iteratively for each new layer, takes about 70 minutes on a 2.00 GHz CPU.
5) We use direct simulation without a priori data transformation except for the use of logarithmic *K* values and orthogonalization. The marginal primary data histogram is approximated by post-processing while keeping the multivariate data conditioning intact, following the approach of Journel & Xu (1994; see also Goovaerts 1997):

$$z_{cc}(u_\alpha) = z(u_\alpha) + \left(\frac{\sigma(u_\alpha)}{\sigma_{max}}\right)^{\varphi} \left(z_c(u_\alpha) - z(u_\alpha)\right) \qquad [8.6]$$

with $z_{cc}(u_\alpha)$ the final corrected value accounting for the data conditioning at location u_α, z_c the correction of the initial z value using rank transformation to obtain the target histogram, σ the kriging standard deviation obtained from all primary and secondary data, and σ_{max} the maximum of the latter. The parameter φ is the correction parameter and allows a balance

between correction at all locations ($\varphi=0$), and no correction at all ($\varphi=\infty$). We use $\varphi=0.05$ after testing different values. Figure 8.7 provides a comparison between the histograms of the simulated data and the target primary data.

The best performing (in terms of RMSE) out of 20 geostatistical realizations was selected for the direct-push-conditioned HUF3D section. Then local optimization was performed for the following parameters: 1) a multiplication factor for the HUF1D K_h values, 2) a multiplication factor for the HUF1D vertical anisotropy values, 3) a multiplication factor for the overall groundwater recharge, and 4) the built area recharge percentage compared to the meadow recharge. The argument for optimising the HUF1D parameters is that we do expect systematic bias due to scaling issues given our 1D approach. We have much more data in 3D for the HUF3D section, so we expect less bias due to upscaling there and do not optimize the marginal distribution of that data; we only vary its spatial distribution. As we do not optimize the marginal distribution of the HUF3D data, we do calibrate the recharge. For obtaining accurate travel times, and deriving the true balance between recharge and the HUF3D data marginal distribution, groundwater age tracing is highly recommended. Starting values were set at 1.5, 1.5, 0.6, and 0.7 respectively, after manual exploration of the parameter space. Automatic calibration was performed with the limited memory quasi-Newton algorithm of Byrd *et al.* (1995) for solving large nonlinear optimization problems with simple bounds on the variables.

To evaluate the model performance, the calibrated reference simulation with homogeneous hydrogeological units is used, as in Rogiers *et al.* (2012b), including the numerical grid and observation adaptations, which have a small but distinguishable effect on the model performance.

8.3. Results and discussion

8.3.1. Calibration of DPT data for *K* estimation

Table 8.2 summarizes the calibration approach and results for the different types of high-resolution direct push data. The resulting calibrated meter-scale hydraulic direct push K values (all K_h) are plotted in Figure 8.8 versus the reliable K_{DPST} results. The coefficient of determination is 0.82 for the HPT data, and 0.74 for the DPIL data. The previous higher value is probably a consequence of the higher resolution of the HPT logs, compared to the ordinary kriging of the lower resolution DPIL logs. Given these high coefficients of determination, we decided to add this calibrated high-resolution direct push data to the primary variable in the geostatistical approach. The leave-one-out cross-validation results in Table 8.2 show that similar accuracy can be expected for the high-resolution hydraulic direct push log sections without corresponding K_{DPST} data.

The point-scale relationship between I_c and the high-resolution borehole core K logs from Rogiers *et al.* (2013b; Chapter 5) is displayed in Figure 8.9A. The coefficient of determination of 0.32 is rather low, which is probably due to the difficult matching of borehole core and CPT logs at the cm scale. Moreover, the lowest K values clearly remain undetected by the I_c parameter due to the support volume of the CPT measurement. The results obtained after calibration with the BE_{Fr} approach, and upscaling to meter-scale values, are shown in Figure 8.9B. Coefficients of determination are 0.43 and 0.75 for respectively K_h and K_v. For the latter, this is similar to the hydraulic direct push data. The leave-one-out cross-validation in Table 8.2 again shows that similar accuracy is expected for the CPT-based K values for the entire dataset. Given the lower coefficient of determination for the CPT-based K_h and the fact that these values are derived from geotechnical characteristics of the subsurface, we decided to treat this data as a secondary variable in the geostatistical framework.

For illustration of the CPT-based K estimates, the upscaled primary and calibrated secondary data at three cored boreholes is shown in Figure 8.10.

8.3.2. Hydraulic conductivity realizations

The non-stationary (along z_{strat}) means for logarithmic K_h and K_v and their PCs, and the correlation coefficients between primary and secondary data (for the PCs) are shown in Figure 8.11. The K means clearly illustrate the presence of the Kasterlee Clay aquitard, mostly between 0 and -5 m stratigraphical depth (Figure 8.11A). The lower aquifer also shows lower K values than the upper aquifer. For the uppermost and lowest HUF layers, the means fall back to the global mean, because no or only very little data is available at these locations. The same

holds for the PC means in Figure 8.11B. The correlation coefficient between primary and secondary data is overall above 0.75, except for an interval around 25 m stratigraphical depth, which corresponds to the coarse white Mol Upper Sands. This is a consequence of the CPT data hardly giving an indication on this coarse high *K* unit (Figure 8.4A,C), as it shows lower Q_{tn} values. Hence, this discrepancy between secondary and primary data is accounted for.

The results of the distance-weighted variography are shown in Figure 8.12. All variogram models consist of a nugget value and two spherical models. A first observation is the high semivariances of the 0 and -10 m stratigraphical depth levels. These are highly influenced by the heterogeneity within the Kasterlee Clay aquitard. A second consistency between all variograms are the low semivariances, especially at short lag distances, of the 20, 30 and 40 m stratigraphical depth locations. These correspond to the Mol Formation units, which are the most homogeneous sediments in the study area. The other variograms fall in between these two extreme cases, both with their total sill and relative nugget values.

The K_h and K_v data for three example HUF layers of the three best performing (in terms of RMSE) out of 20 realizations are shown in Figure 8.13. The contrast between the upper aquifer (HUF layer 25), Kasterlee Clay aquitard (layer 43) and lower aquifer (layer 50) is very clear. The highest *K* values are located within the upper aquifer, where only very little anisotropy seems to be present. The Kasterlee Clay aquitard clearly has low *K* values, but the K_h data can easily be as large as that of the aquifer units. This results in a large range of vertical anisotropy values. The lower aquifer K_h values are clearly somewhat lower than that of the upper aquifer, but the K_v values are clearly lower, which also results in a non-negligible vertical anisotropy.

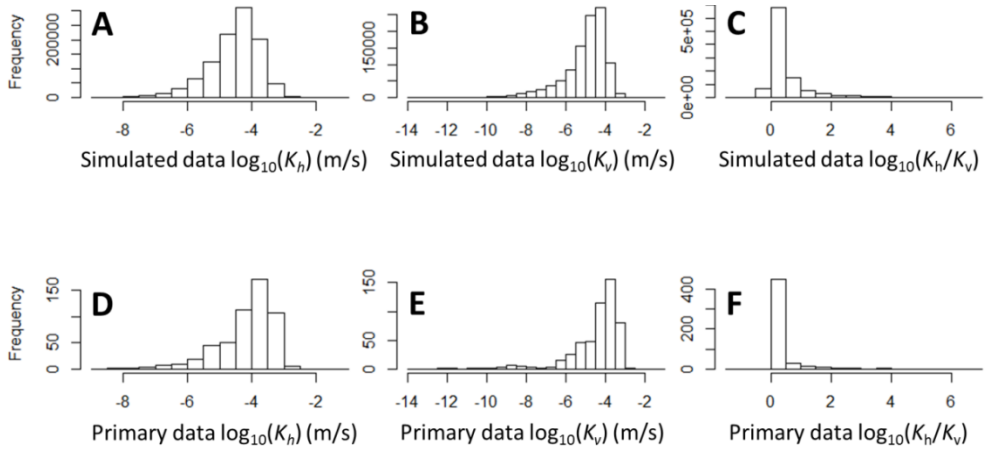

Figure 8.7: Comparison between simulated (A, B, C) and target primary (D, E, F) data histograms for K_h (A, D), K_v (B, E) and vertical anisotropy (C, F).

Table 8.2: Overview of the high-resolution direct push data calibration.

Direct push technology	Calibration data	# (ups)	Calibration method	Calibration		Cross-validation		Log10(Kups)	
				R^2	RMSE	R^2	RMSE	min	Max
DPIL	DPST	13	LM, OK, Upscaling	0.74	0.18	0.68	0.19	-4.1	-3.2
HPT	DPST	10	LM, Upscaling	0.82	0.16	0.74	0.19	-4.1	-3.2
CPT	OCK constant head & air permea-meter tests	201	LM, BE_{Fr} correction, LM, Upscaling	**Kh**					
				0.43	0.79	0.29	0.9	-7.5	-2.8
				Kv					
				0.75	0.95	0.68	1.07	-12	-3.6

OCK = ordinary co-kriging ; ups = upscaled data; OK = ordinary kriging; LM = linear model; BE_{Fr} = Friction ratio boundary energy

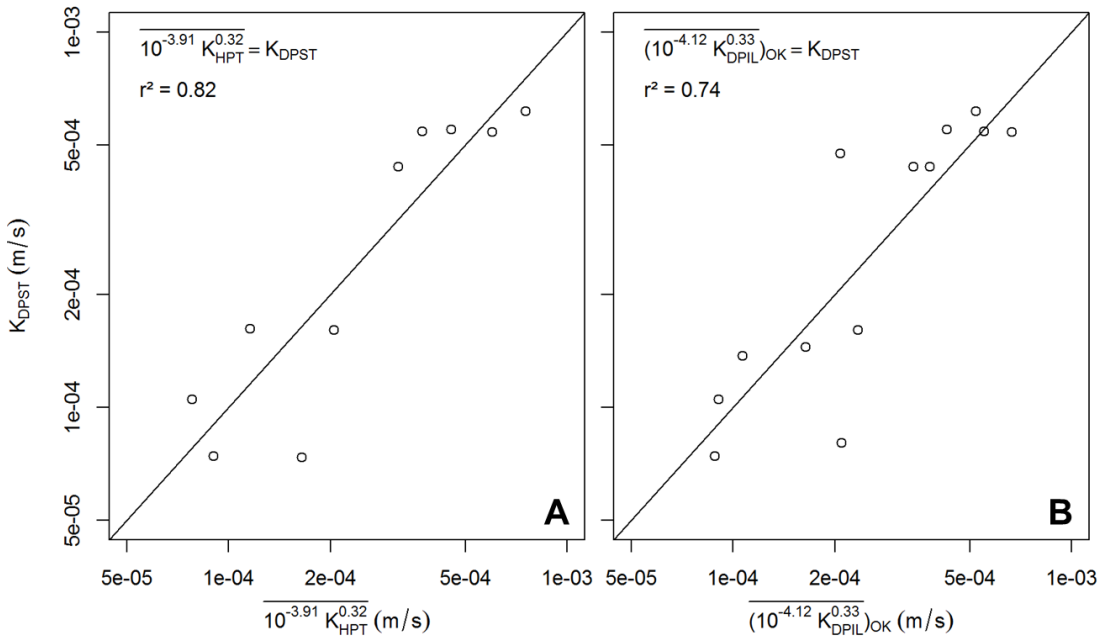

Figure 8.8: Calibration results for upscaled A) K_{HPT} and B) K_{DPIL} data, using the absolute K_{DPST} data. OK denotes the ordinary kriging that is performed for the DPIL data, to get a better DPST screen support estimate.

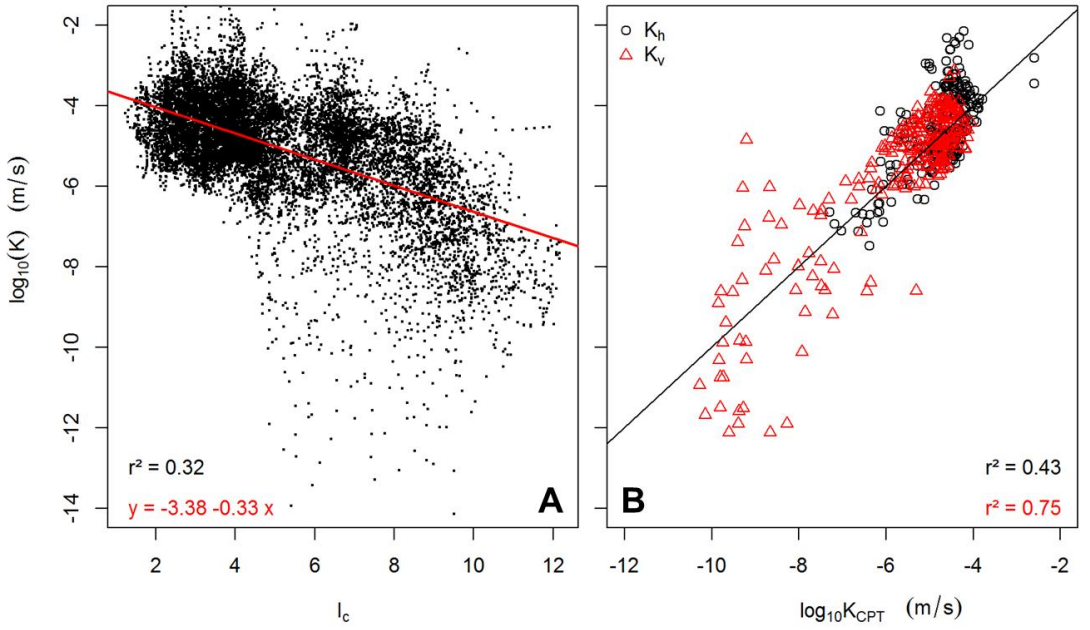

Figure 8.9: A) Scatterplot of the high-resolution borehole core K logs and the corresponding soil behaviour type index (I_c). B) Scatter plot of calibrated $K_{h,CPT}$ and $K_{v,CPT}$ versus upscaled high-resolution K logs for the cored boreholes, obtained by Rogiers *et al.* (2013b; Chapter 5).

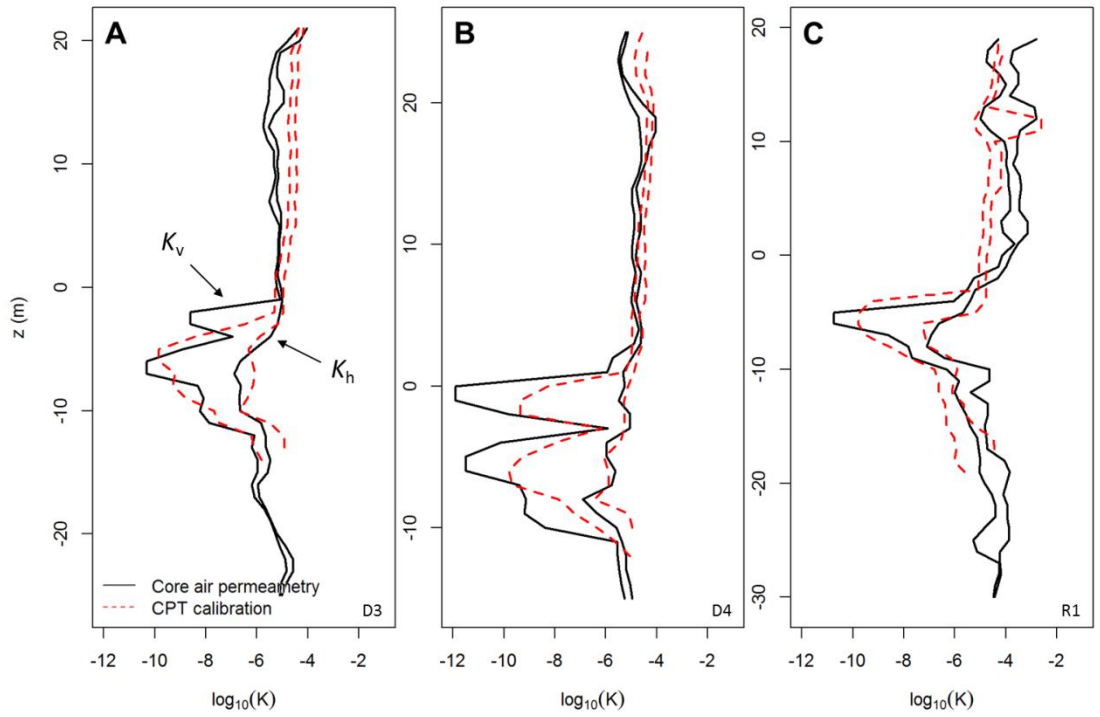

Figure 8.10: Comparison between upscaled borehole core *K* logs and CPT-calibrated K_h (left) and K_v (right) logs for the Dessel-3 (A), Dessel-4 (B), and Retie-1 (C) borehole cores.

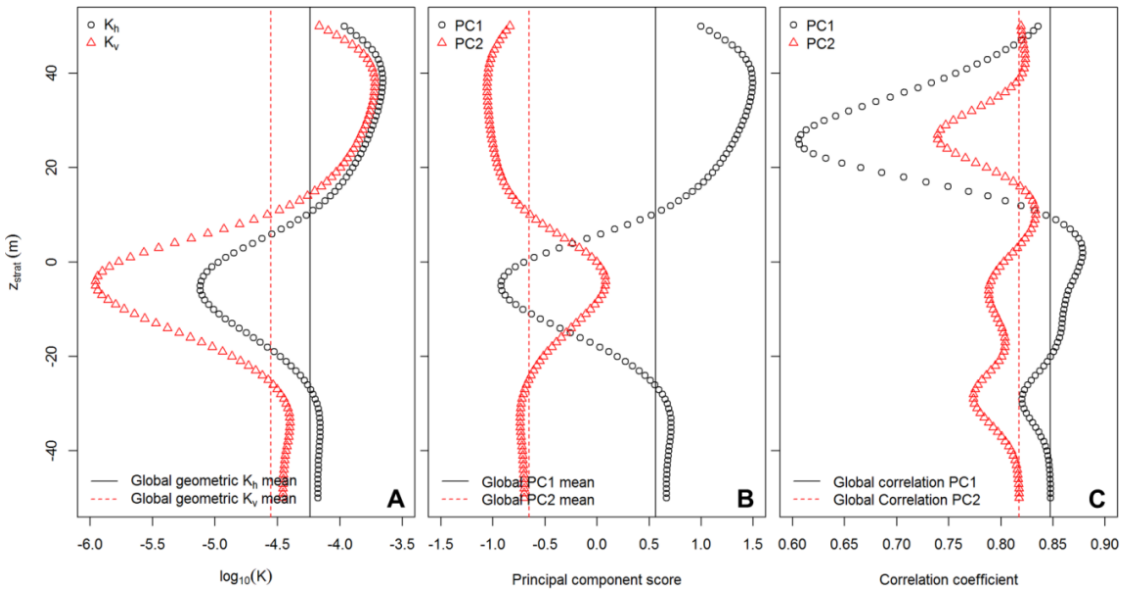

Figure 8.11: A) Non-stationary geometric K means along z_{strat}. B) Non-stationary principal component (PC1 and PC2) means along z_{strat}. C) Non-stationary correlation coefficients between primary and secondary data PCs.

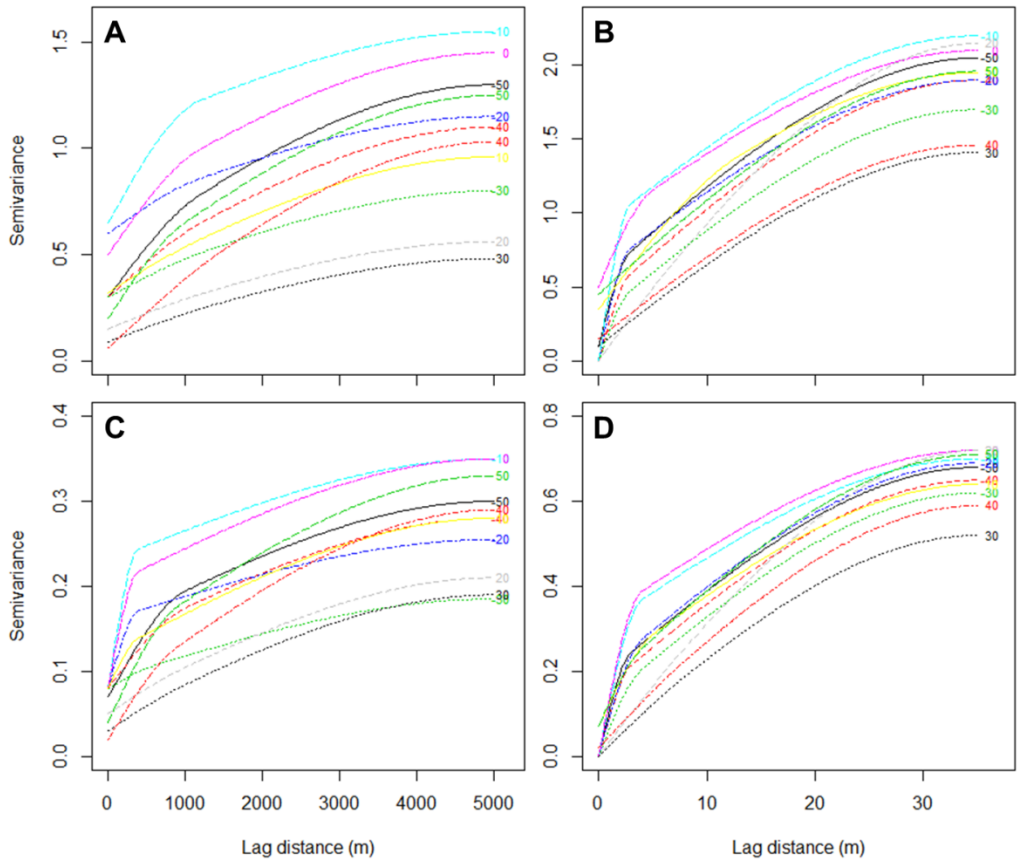

Figure 8.12: Overview of the distance-weighted variogram models for PC1 (A, B) and PC2 (C, D). Both horizontal (A, C) and vertical (B, D) variograms are shown. Different colour codes are for different stratigraphical depths between -50 and 50 m (with respect to the top of the aquitard), at 10 m intervals.

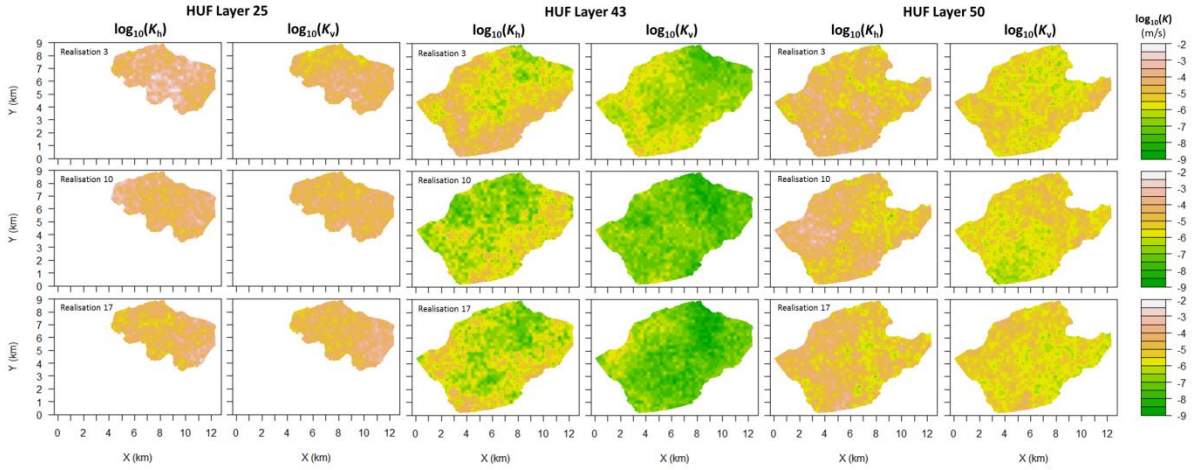

Figure 8.13: Example K_h and K_v fields of the three best performing realizations out of 20. HUF layer 25 represents the upper aquifer, layer 43 the Kasterlee Clay aquitard, and layer 50 the lower aquifer.

For the Kasterlee Clay aquitard, several of the realizations seem to suggest the presence of a SW-NE oriented zone with the lowest *K* values. This could be related to the existence of channels in the same direction, linked to the fluvio-marine or estuarine depositional environment

8.3.3. Groundwater flow model performance testing

The locally optimized, best performing realization out of a set of 20 realizations of the HUF3D *K* parameters was selected for comparison with the reference case. Figure 8.14 shows the model predictions, or simulated equivalents, versus the observed head values, as well as the vertical head differences (inset graph) that were added to the objective function. The total RMSE decreased from 0.37 to 0.35 with the model updates presented in this paper. The improvement is most easily noticeable for the vertical head differences, which reflects directly the more appropriate combination of recharge and *K* values in the model update. Many of the simulated head values are also much closer to the observations, but a few anomalies seem to be present in the updated model, with a misfit of more than one meter. These are influenced strongly by local characteristics of the *K* field, as they are located in the very sensitive southern part of the model (Gedeon & Mallants 2012), where the Kasterlee Clay aquitard is at its shallowest position. The use of a larger set of realizations, or the optimization of the spatially distributed parameters, honouring the geostatistical model, should therefore be investigated.

The RMSE values of all 20 realizations are shown in Figure 8.15. This also indicates the need for optimization of the spatially distributed parameters, as random realizations, even honouring all primary and secondary data, certainly do not always reproduce the groundwater heads and head differences to a higher degree than the reference case.

Given the fact that this highly data-conditioned model performs better than the homogeneous reference model in certain realizations, an in depth-analysis of the obtained flow fields might reveal important aspects on the hydrogeological system, which can be in contrast with the previous conceptual hydrogeological model. Further analyses to quantify the inter-aquifer connectivity and groundwater residence times are highly recommended.

8.3.4. Solute transport

For illustrating the consequences on contaminant transport of accounting for aquifer and aquitard heterogeneity using direct push data, we performed two conservative tracer simulations. The tracer is inserted by mass loading at the locations of the two tumuli of the future cAt project (ONDRAF/NIRAS, 2010) disposal facility (solute source in Figure 8.16). A total flux of 1E6 Bq/y is specified, and divided equally amongst the two tumuli. A steady-state MT3DMS (Zheng & Wang 1999) transport simulation is performed for the reference and the updated model flow fields using the LMT6 package (Zheng *et al.*, 2001), with a steady-state finite difference advection sheme. These results are directly comparable to the simulations reported by Gedeon *et al.* (2011), but further investigations are required given the considerable influence of numerical dispersion with this advection solution. Maps and plumes of the resulting total activity within the Neogene aquifer are shown in Figure 8.16. The spatial extent of the plumes, especially for the left tumulus, is very different in the two cases. These results clearly illustrate the importance of accounting for subsurface heterogeneity, and the usefulness of direct push data for aquifer characterization.

8.4. Conclusions

Direct push technologies are very much suited for aquifer characterization from the local to the regional scale. Moreover, we have illustrated that standard cone penetration test data can be used to estimate *K* if calibration data is available, allowing to map aquifer and aquitard properties in 3D. The use of hydraulic direct push tools is more time consuming, but a higher accuracy can be obtained with similar data resolution. Both types of direct push data were integrated together with available borehole data in a regional groundwater flow model. We considered all hydraulically obtained data as the primary variable, while the geotechnical data was considered as a secondary variable. A non-stationary geostatistical approach was used to condition the model parameter realizations. After local optimization, and the selection of the best out of 20 realizations, a considerable

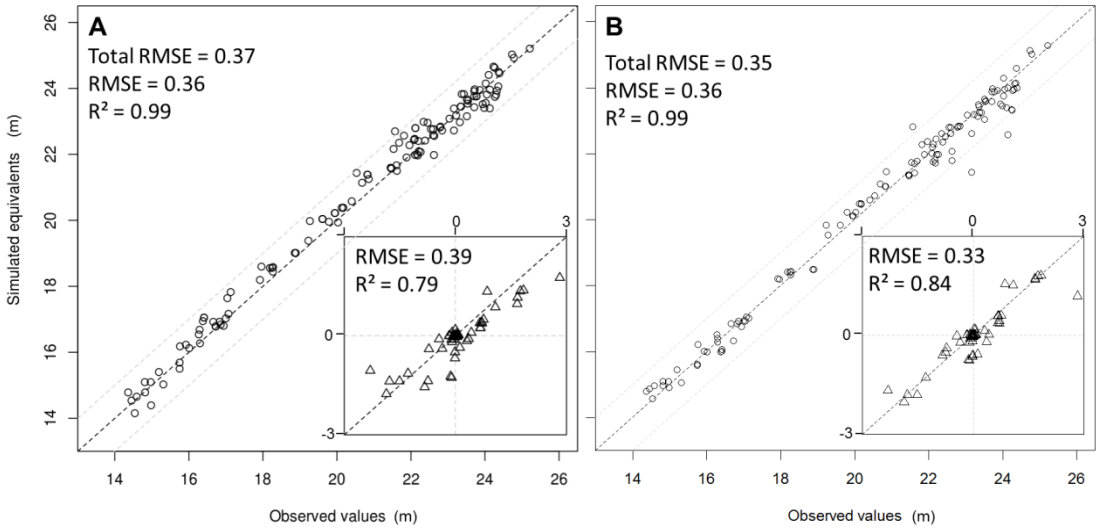

Figure 8.14: Predicted versus observed heads for A) the reference case with homogeneous hydrogeological layers, and B) the best performing direct push-conditioned realization. The dashed lines represent a residual of 1 m. The inset shows the vertical head differences used in the objective function.

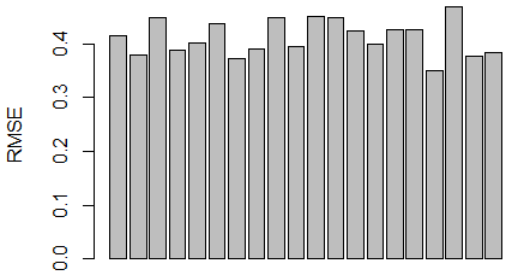

Figure 8.15: RMSE values for the 20 different geostatistical realizations, with the same global parameters.

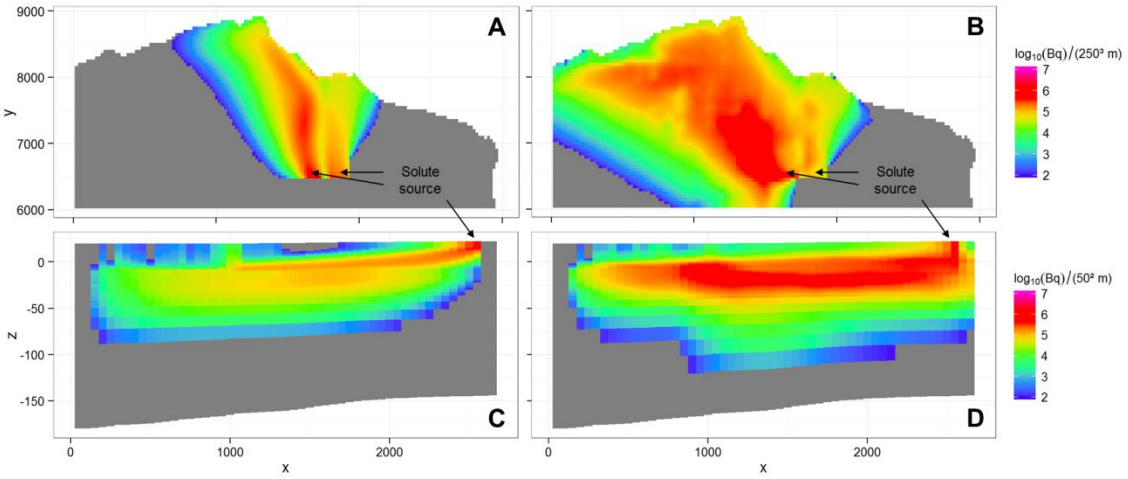

Figure 8.16: Maps (A, B) and profiles (C, D) of total activity within the Neogene aquifer, based on the steady-state solute plume for the homogeneous reference (A, C) and the heterogeneous improved (B, D) realization of the groundwater flow model.

increase in model performance was obtained, compared to the reference case. An example transport simulation clearly showed the impact that subsurface heterogeneity can have on contaminant distributions, and that it should be accounted for. Therefore we recommend using effective direct push technologies for shallow aquifer heterogeneity characterization.

As the CPT data calibration in this paper was performed using available cored borehole data, future research should focus on omitting the need for such data, and replacing it with hydraulic direct push data. A few well-chosen hydraulic direct push test sites could then complement a regional CPT campaign, and allow for high-resolution, high-detail multi-scale aquifer characterization without the need for expensive drilling.

Concerning the presented case study, we conclude that the hydraulic direct push data is compatible with the cored boreholes, and clearly images a high *K* sand layer that was overlooked up to know in the hydrogeological conceptual model. The CPT data calibration however does not really capture this high *K* zone (coarse white Upper Mol Sand in Figure 8.3), which indicates that calibration data for the entire lithostratigraphical column is preferred when using CPTs. The non-stationary geostatistical approach allowed however to account for this discrepancy between primary and secondary data.

The improvement in groundwater flow model performance, and the reduced uncertainty of the contaminant transport simulation should be evaluated stochastically in future work to provide feedback to the ONDRAF/NIRAS cAt surface disposal program. This should include the use of realistic dispersion values, that may be based on the direct push data or *K* realizations, and an analysis of the ratio between numerical and physical dispersion in the simulations.

A similar approach could be used for characterization of effective porosity, which is important for transient solute transport. Measurements of this porosity are however also not straightforward, and the uncertainty on this parameter is much less than that on *K*, which varies across different orders of magnitude, so it is less important.

References

Anderman, E.R., Hill, M.C., 2000. MODFLOW-2000, the U.S. Geological Survey Modular Ground-water model – documentation of the hydrogeologic-unit flow (HUF) package. USGS Open file report 00-342, Denver, Colorado

Battle-Aguilar, J., Brouyere, S., Dassargues, A., Morasch, B., Hunkeler, D., Hohener, P., Diels, L., Vanbroekhoeven, K., Seuntjens, P., Halen, H., 2009. Benzene dispersion and natural attenuation in an alluvial aquifer with strong interactions with surface water, *Journal of Hydrology*, **361**: 305-317.

Bhattacharya, B., Solomatine, D.P., 2005. Machine learning in soil classification. *Proceedings of International Joint Conference on Neural Networks* **19**: 2694 – 2699.

Beerten, K., Wemaere, I., Gedeon, M., Labat, S., Rogiers, B., Mallants, D., Salah, S., Leterme, B., 2010. Geological, hydrogeological and hydrological data for the Dessel disposal site. Project near surface disposal of category A waste at Dessel. STB-SIE(HYD) - Version 1, NIROND-TR 2009-05 E (p. 261).

Bierkens, M.F.P., 2005. Designing a monitoring network for detecting groundwater pollution with stochastic simulation and a cost model. *Stochastic Environmental Research and Risk Assessment* **20**(5): 335–351. doi:10.1007/s00477-005-0025-2

Butler, J.J., 2002. A simple correction for slug tests in small-diameter wells. *Ground Water* **40**(3): 303–308.

Butler, J.J., Healey, J.M., McCall, G.W., Garnett, E.J., Loheide II, S.P., 2002. Hydraulic tests with direct-push equipment. *Ground Water* **40**(1): 25–36.

Butler, J.J., Dietrich, P., Wittig, V., Christy T., 2007. Characterizing hydraulic conductivity with the direct-push permeameter. *Ground Water* **45(**4): 409–419.

Byrd, R.H., Lu, P., Nocedal, J., Zhu, C., 1995. A limited memory algorithm for bound constrained optimization. *SIAM J. Scientific Computing* **16**: 1190–1208.

Chai, J., Agung, P., Hino, T., Igaya, Y., Carter, J., 2011. Estimating hydraulic conductivity from piezocone soundings. *Géotechnique* **61**(8) : 699–708. doi:10.1680/geot.10.P.009

Chai, J., Sheng, D., Carter, J.P., Zhu, H., 2012. Coefficient of consolidation from non-standard piezocone dissipation curves. *Computers and Geotechnics* **41**: 13–22. doi:10.1016/j.compgeo.2011.11.005

Christian, J., Van Der Sluijs, J.P., Brown, J., Van Der Keur, P., 2006. A framework for dealing with uncertainty due to model structure error. *Advances in Water Resources* **29**: 1586–1597. doi:10.1016/j.advwatres.2005.11.013

Christian, J., Van Der Sluijs, J.P., Lajer, A., Vanrolleghem, P.A., 2007. Uncertainty in the

environmental modelling process - A framework and guidance. *Water* **22**: 1543–1556. doi:10.1016/j.envsoft.2007.02.004

Dietrich, P., Leven, C., 2006. Direct Push-Technologies. In: Kirsch, R. (Ed.), Groundwater Geophysics. Springer, Berlin, pp. 321-340.

Dietrich, P., Butler, J.J., Faiss, K., 2008. A rapid method for hydraulic profiling in unconsolidated formations. *Ground water* **46**(2): 323–8.

Elsworth, D., Lee, D.S., 2007. Limits in determining permeability from on-the-fly uCPT sounding. *Géotechnique* **57**(8): 679–685. doi:10.1680/geot.2007.57.8.679

Flach, G.P., Harris, M.K., Smits, A.D., Syms, F.H., 2005. Modeling aquifer heterogeneity using cone penetration testing data and stochastic upscaling methods. *Environmental Geosciences* **12**(1): 1–15.

Gedeon, M., 2008. Neogene Aquifer Model, SCK•CEN-ER-48 (p. 110). Mol.

Gedeon, M., Mallants, D., Vandersteen, K., Rogiers, B., Laloy, E., 2011. Hydrogeological modelling of the Dessel site - Overview report (p. 227).

Gedeon, M., Mallants, D., 2012. Sensitivity Analysis of a Combined Groundwater Flow and Solute Transport Model Using Local-Grid Refinement: A Case Study. *Mathematical Geosciences* **44**(7): 881–899. doi:10.1007/s11004-012-9416-3

GEOPROBE, 2007. GEOPROBE hydraulic profiling tool (HPT) systems: Standard operating procedure. Technical Bulletin MK3137, GEOPROBE.

Hammond, G.E., Lichtner, P.C., 2010. Field-scale model for the natural attenuation of uranium at the Hanford 300 Area using high-performance computing. *Water Resources Research* **46**(9): 1–31. doi:10.1029/2009WR008819

Harbaugh, A.W., 2005. MODFLOW-2005, The US Geological Survey modular ground-water model—the Ground-Water Flow Process: US Geological Survey Techniques and Methods 6-A16, USGS, 2005.

Harris, P., Charlton, M., Fotheringham, A., 2010. Moving window kriging with geographically weighted var- iograms. *Stoch Environ Res Risk Assess* **24**: 1193–1209. doi:10.1007/s00477-010-0391-2

Huysmans, M., Madarász, T., Dassargues, A., 2006. Risk assessment of groundwater pollution using sensitivity analysis and a worst-case scenario analysis. *Environmental Geology* **50**(2): 180–193. doi:10.1007/s00254-006-0197-1

Journel, A.G., Xu, W., 1994. Posterior Identification of histograms Conditional to Local Data. *Mathematical Geology* **22**: 323-359.

Köber, R., Hornbruch, G., Leven, C., Tischer, L., Grossmann, J., Dietrich, P., Weiss, H., Dahmke, A., 2009. Evaluation of combined direct-push methods used for aquifer model generation. *Ground Water* **47**(4): 536–46. doi:10.1111/j.1745-6584.2009.00554.x

Labat, S., Gedeon, M., Beerten, K., Maes, T., 2011. Dessel-5 borehole: technical aspects and hydrogeological investigations. SCK•CEN-ER-151 (p. 39).

Laga, P., Louwye, S., Geets, S., 2001. Paleogene and Neogene lithostratigraphic units (Belgium). *Geologica Belgica* **4**(1-2): 135–152.

Lee, D.S., Elsworth, D., Hryciw, R., 2008. Hydraulic Conductivity Measurement from On-the-Fly uCPT Sounding and from VisCPT. *Journal of Geotechnical and Geoenvironmental Engineering* **134**(12): 1720-1729. doi:10.1061/(ASCE)1090-0241(2008)134:12(1720)

Lessoff, S.C., Schneidewind, U., Leven, C., Blum, P., Dietrich, P., Dagan, G., 2010. Spatial characterization of the hydraulic conductivity using direct-push injection logging. *Water Resources Research* **46**(12): 1–9.

Leterme, B., Mallants, D., 2012. Simulation of evapotranspiration and groundwater recharge in the nete catchment accounting for different land cover types and for present and future climate conditions. Project near surface disposal of category A waste at Dessel. External report of the Belgian Nuclear Research Centre. SCK•CEN-ER-192.

Liu, G., Butler, J.J., Bohling, G.C., Reboulet, E., Knobbe, S., Hyndman, D.W., 2009. A new method for high-resolution characterization of hydraulic conductivity. *Water Resources Research* **45**(8): 1-6.

Louwye, S., De Schepper, S., Laga, P., Vandenberghe, N., 2007. The Upper Miocene of the southern North Sea Basin (northern Belgium): a palaeoenvironmental and stratigraphical reconstruction using dinoflagellate cysts. *Geological Magazine* **144**(1): 33-52.

Louwye, S., Laga, P., 2008. Dinoflagellate cyst stratigraphy and palaeoenvironment of the marginal marine Middle and Upper Miocene of the eastern Campine area, northern Belgium (southern North Sea Basin). *Geological Journal* **43**(1): 75–94.

Louwye, S., De Schepper, S., 2010. The Miocene–Pliocene hiatus in the southern North Sea Basin (northern Belgium) revealed by dinoflagellate cysts. *Geological Magazine* **147**(05): 760-776.

Lunne, T., Robertson, P.K., Powell, J.J.M., 1997. Cone Penetration Testing in Geotechnical Practice. Blackie Academic/Routledge Publishing, New York.

McCall, W., Nielsen, D.M., Farrington, S.P., Christy, T.M., 2006. Use of Direct-Push Technologies in Environmental Site Characterization and Ground-Water Monitoring. In: Nielsen, D.M. (Ed.), Practical Handbook of Environmental Site Characterization and Ground-Water Monitoring. CRC Press Taylor and Francis Group, Boca Raton, pp. 345-471.

McCall, W., Christy, T.M., Christopherson, T., Issacs, H., 2009. Application of Direct Push Methods to Investigate Uranium Distribution in an Alluvial Aquifer. *Ground Water Monitoring & Remediation* **29**(4): 65-76.

McLennan, J.A., 2002. The Effect of the Simulation Path in Sequential Gaussian Simulation. Centre for Computational Geostatistics report 2002-54 (p. 16).

Nelder, J.A., Mead, R., 1965. A simplex algorithm for function minimization. *Computer Journal* **7**: 308–313.

ONDRAF/NIRAS, 2010. Het cAt-project in Dessel. Een langetermijnoplossing voor het Belgische categorie A-afval. Retrieved from http://www.niras-cat.be/downloads/cAt_masterplan_NL_LOW.pdf on 07-12-2011.

Patyn, J., Ledoux, E., Bonne, A., 1989. Geohydrological research in relation to radioactive waste disposal in an argillaceous formation. *Journal of Hydrology* **109**(3-4): 267–285.

Pebesma, E.J., 2004. Multivariable geostatistics in S: the gstat package. *Computers & Geosciences* **30**: 683-691.

Robertson, P.K., 1990. Soil classification using the cone penetration test. *Canadian Geotechnical Journal* **27**(1): 151-158.

Robertson, P.K., Wride, C.E., 1998. Evaluating cyclic liquefaction potential using the cone penetration test, *Canadian Geotechnical Journal* **35**: 442 – 459.

Robertson, P., 2010. Estimating in-situ soil permeability from CPT & CPTu. CPT'10, 2nd international symposium on cone penetration testing **2**: 51.

Rojas, R., Feyen, L., Dassargues, A., 2008. Conceptual model uncertainty in groundwater modeling: Combining generalized likelihood uncertainty estimation and Bayesian model averaging. *Water Resources Research* **44**(12): 1–16. doi:10.1029/2008WR006908

Rogiers, B., Schiltz, M., Beerten, K., Gedeon, M., Mallants, D., Batelaan, O., Dassargues, A., Huysmans, M., 2010. Groundwater model parameter identification using a combination of cone penetration tests and borehole data. Proceedings of the IAHR International Groundwater Symposium 2010, Valencia, 22-24 September 2010, 19 pp.

Rogiers, B., Mallants, D., Batelaan, O., Gedeon, M., Huysmans, M., Dassargues, A., 2011. Site-specific soil classification from cone penetration tests and borehole data: a multivariate statistical analysis. NovCare 2011, Cape Cod, 9-11 May 2011.

Rogiers, B., Mallants, D., Batelaan, O., Gedeon, M., Huysmans, M., Dassargues, A., 2012a. Estimation of hydraulic conductivity and its uncertainty from grain-size data using GLUE and artificial neural networks. *Mathematical Geosciences* **44**(6): 739–763.

Rogiers, B., Mallants, D., Batelaan, O., Gedeon, M., Huysmans, M., Dassargues, A., 2012b. The usefulness of CPTs for deterministic, spatially heterogeneous, large-scale aquitard parameterisation. In Oswald, S.E., Kolditz, O., Attinger, S. (Eds.), *Models - Repositories of Knowledge, IAHS Publ. 355* (pp. 41–47).

Rogiers, B., Winters, P., Huysmans, M., Beerten, K., Mallants, D., Gedeon, M., Batelaan, O., Dassargues, A., 2012c. Centimeter-scale secondary information on hydraulic conductivity using a hand-held air permeameter on borehole cores. *Geophysical Research Abstracts* **14**: EGU2012-1794-1. EGU General Assembly 2012, Vienna, 22-27 April 2012.

Rogiers, B., Beerten, K., Smeekens, T., Mallants, D., Gedeon, M., Huysmans, M., Batelaan, O., Dassargues, A., 2013a. Derivation of flow and transport parameters from outcropping sediments of the Neogene aquifer, Belgium. *Geologica Belgica* **16**(3): 129-147.

Rogiers, B., Winters, P., Huysmans, M., Beerten, K., Mallants, D., Gedeon, M., Batelaan, O., Dassargues, A., 2013b. High resolution hydraulic conductivity logging of borehole cores using air permeability measurements. *Hydrogeology Journal*, submitted.

Rogiers, B., Mallants, D., Batelaan, O., Gedeon, M., Huysmans, M., Dassargues, A., 2013c. High-resolution hydrostratigraphic characterization of a heterogeneous sedimentary aquifer using cone penetration data. *Water Resources Research*, submitted.

Saito, H., Goovaerts, P., 2003. Selective remediation of contaminated sites using a two-level multiphase strategy and geostatistics. *Environmental science & technology* **37**(9): 1912–8.

Schiltz, M., 2008. Lithological and Stratigraphical interpretation by means of cone penetration tests (CPT's) in the Dessel-Kasterlee-Geel-Mol area. Bvba SAMSUFFIT Geoservices, Fieldsurvey cAt 2008.

Schiltz, M., 2010. Lithological and Stratigraphical interpretation of cone penetration tests (CPT's) executed for the first tumulus at the disposal site in Dessel and in the Dessel-Kasterlee-Geel-Mol area. Bvba SAMSUFFIT Geoservices, Fieldsurvey cAt 2010.

Selroos, J.-O., Painter, S.L., 2012. Effect of transport-pathway simplifications on projected releases of radionuclides from a nuclear waste repository (Sweden). *Hydrogeology Journal* **20**:1467–1481. doi:10.1007/s10040-012-0888-5

Šimůnek, J., Sejna, M., van Genuchten, M.T., 2005. HYDRUS-1D, version 4.14, code for simulating the one-dimensional movement of water, heat, and multiple solutes in variably saturated porous

media, Tech. rep., University of California Riverside, 2005.

Teh, C., Houlsby, G., 1991. An analytical study of the cone penetration test in clay. *Geotechnique* **41**(1): 17-34.

Tillmann, A., Englert, A., Nyari, Z., Fejes, I., Vanderborght, J., Vereecken, H., 2008. Characterization of subsoil heterogeneity, estimation of grain size distribution and hydraulic conductivity at the Krauthausen test site using Cone Penetration Test. *Journal of contaminant hydrology* **95**(1-2): 57–75.

Tinter, M., 2012. Vergleich direkter und indirekter Verfahren zur Bestimmung des Durchlässigkeitsbeiwertes an klastischen Sedimenten der Mol-Formation am Standort Dessel in Belgien. Masterarbeit. Universität Leipzig.

Vienken, T., Tinter, M., Rogiers, B., Leven, C., Dietrich, P., 2012. Evaluation of field methods for vertical high resolution aquifer characterization. Abstract ID 1494473, AGU Fall Meeting, San Francisco, USA, 3-7 December 2012.

Wouters, L., Schiltz, M., 2012. Overview of the field investigations in and around the nuclear site of Mol-Dessel, NIROND-TR 2011-42.

Mariethoz, G., Renard, P., Caers, J., 2009. Bayesian inverse problem and optimization with Iterative Spatial Resampling. *Water Resources Research* **46**(11): 1–17. doi:10.1029/2010WR009274

Nilsson, B., Højberg, A., Refsgaard, J., Troldborg, L., 2007. Uncertainty in geological and hydrogeological data. *Hydrology and Earth System Sciences* **11**, 1551–1561.

Machuca-Mory, D.F., Deutsch, C.V., 2012. Non-stationary Geostatistical Modeling Based on Distance Weighted Statistics and Distributions. *Mathematical Geosciences* **45**(1): 31–48. doi:10.1007/s11004-012-9428-z

Van Baars, S., Van De Graaf, H.C., 2007. Determination of Organic Soil Permeability Using The Piezocone Dissipation Test. *Environmental and Engineering Geoscience* **13**(3): 197–203. doi:10.2113/gseegeosci.13.3.197

Yu, L., Rogiers, B., Gedeon, M., Marivoet, J., De Craen, M., Mallants, D., 2013. A critical review of laboratory and in-situ hydraulic conductivity measurements for the Boom Clay in Belgium. *Applied Clay Science* **75-76**: 1–12. doi:10.1016/j.clay.2013.02.018

Zheng, C., Wang, P.P., 1999. MT3DMS: A Modular Three-Dimensional Multispecies transport Model for Simulation of Advection, Dispersion, and Chemical Reactions of Contaminants in Groundwater Systems; Documentation and User's Guide. US Army Corps of Engineers, Contract report SERDP-99-1, December 1999.

Zheng, C., Hill, M.C., Hsieh, P.A., 2001. MODFLOW-2000, The U.S. Geological Survey Modular Ground-Water Model—User Guide to the LMT6 Package, the Linkage with MT3DMS for Multi-Species Mass Transport Modeling, U.S. Geological Survey Open File Report 01-82, Denver, Colorado, 2001.

Chapter 9

Large-scale stochastic inversion using non-stationary variograms for transport uncertainty assessment of a near surface radioactive waste disposal

Based on Rogiers B, Laloy E, Gedeon M, Dassargues A, Huysmans M, Batelaan O, Mallants D. Large-scale stochastic inversion using non-stationary variograms for transport uncertainty assessment of a near surface radioactive waste disposal. In preparation.

Abstract

Uncertainty quantification is very much needed to support decision making related to *e.g.* environmental impact assessment for waste disposal sites. A probabilistic result provides a much stronger basis for decision making compared to a single deterministic outcome. Accurate posterior exploration of high-dimensional and CPU-intensive models, which are often used for environmental impact assessment, is however a challenging task. To quantify the uncertainty associated with solute transport in the framework of a near surface radioactive waste disposal in Mol/Dessel, Belgium, we investigate combining the adaptive Metropolis McMC algorithm for updating the global model parameters, and adaptive spatial resampling for updating of the spatially distributed model parameters, by block sampling. The forward model used is a groundwater flow model conditioned on borehole and direct push data, that accounts for non-stationary heterogeneity in hydraulic conductivity. The obtained flow solutions are used for solute transport simulations, and the results are compared with a different groundwater flow model parameterization, that makes use of homogeneous hydrogeological layers. Moreover, a number of simulations is performed to assess the effect of a realistic dispersivity. The results suggest that the approach with homogeneous layers results in conservative estimates of the dilution rate, which has a very specific definition. The few simulations performed for assessing the effect of realistic dispersivity show that a decrease of the dilution rate might be possible as well.

9.1. Introduction

Uncertainty quantification (UQ) of environmental models has been the subject of intensive research in the last decade (*e.g.* Beven & Freer 2001; Rojas *et al.*, 2008; Vrugt *et al.*, 2009; Montanari *et al.*, 2009) as the model prediction uncertainties are at least as important as the predictions themselves (Refsgaard *et al.*, 2007). In the context of subsurface flow and transport modelling, performing UQ typically means providing an ensemble of model realizations that accurately span the range of possible hydrogeological interpretations, while simultaneously being consistent with the available data and prior information. This is usually done by stochastic inverse modelling in which models/parameters are viewed as probabilistic quantities obeying Bayes law (see, e.g, Robert and Casella (2004) for a good introduction to Bayesian statistics). In such case, the so derived set of solutions is called a posterior model/parameter distribution.

For low and moderately high-dimensional nonlinear inverse problems, different posterior inference methods are available (Haario *et al.*, 2006; ter Braak, 2006; ter Braak and Vrugt, 2008; Vrugt et al., 2008a; Cui et al., 2011; Laloy and Vrugt, 2012). Though considerable progress has been made by using dimensionality

reduction techniques (*e.g.* Tonkin & Doherty 2009; Winton *et al.*, 2011; Laloy *et al.*, 2013; He *et al.*, 2013) and surrogate models (*e.g.* Razavi *et al.*, 2012; Laloy *et al.*, 2013; Kourakos & Mantoglou 2013), accurate posterior exploration of high-dimensional and CPU-intensive models however often remains out of computational reach.

In the formal Bayesian (including Markov chain Monte Carlo -McMC- sampling) versus informal general likelihood estimation (GLUE) debate (*e.g.* Beven & Freer 2001; Stedinger *et al.*, 2008; Vrugt *et al.*, 2008b, Nott *et al.*, 2012), the first approach seems to be less subject to criticism. Besides its proper statistical foundation, one of the advantages of McMC sampling compared to GLUE is the optimization-driven nature of a Markov chain. This makes sampling of consistent solutions by state-of-the-art McMC simulation much more efficient than with the pure brute-force random sampling commonly used with GLUE. The Adaptive Metropolis (AM) algorithm by Haario *et al.* (2001), will be used in this paper.

For sampling spatially distributed variables that honor certain geostatistical properties, different specific tools are available such as pilot points (Doherty 2003) and anchored distributions (Rubin *et al.*, 2010), but these require the arbitrary location of such pilot points and anchors. Moreover, these approaches become increasingly complex in a multi-variate non-stationary geostatistical framework, and the use of regularization is less efficient than forward conditioning of the spatial random fields (because all realizations honour the data). To incorporate spatially distributed parameters using any kind of geostatistical simulation in a McMC framework, Mariethoz *et al.* (2010) proposed the iterative spatial resampling (ISR) method. This method introduces dependence between the proposal geostatistical realization and the previous realization by sampling a certain percentage φ of points from that previous realization. The geostatistical simulation method is then used to simulate the proposal realization, while conditioning on the samples from the previous realization. In this way, a chain of dependant realizations can be created for applying sampling or optimization strategies. To increase the efficiency of this method, Jeong *et al.* (2012) showed that adaptive spatial resampling (ASR) by adjusting the probability to select points for forward conditioning of the proposal realization based on the spatially distributed model misfit at each step of the chain, can increase the efficiency of the method.

When a geostatistical simulation takes a comparable amount of CPU-time than the forward model, it is interesting to decouple the spatially distributed model parameters from the global model parameters. Such an approach would prevent rejecting a good performing realization due to bad performing global parameter proposals within the McMC chain. Such decomposition of the proposal distribution into blocks is typically known as McMC block sampling (Roberts and Casella, 2004).

In this chapter, we try to quantify the uncertainty associated with solute transport simulation in the framework of a near surface radioactive waste disposal in Mol/Dessel, Belgium (ONDRAF/NIRAS 2010), by combining the AM algorithm for updating the global model parameters, and ASR for the separate update of the spatially distributed model parameters. This is based on the borehole and direct push data-conditioned groundwater flow model developed by Rogiers *et al.* (2013c; Chapter 8) in MODFLOW-2005 (Harbaugh 2005), that considers a heterogeneous parameterization of the subsurface, and the reference model described by Gedeon *et al.* (2011) and Gedeon & Mallants (2012) which uses homogeneous hydrogeological layers.

First, a sensitivity analysis of the global parameters of the flow model is presented, to select the parameters relevant for the stochastic simulations. Next, the methods for the uncertainty quantification are provided, and the solute transport model and the used indicators of the solute plume are discussed in detail. Finally, the results on the uncertainty estimation withh the flow model are provided, and the transport outcome is discussed using the previously defined indicators.

9.2. Methods

9.2.1. Sensitivity analysis of the global parameters of the flow model

To decide on the relevant parameters that should be varied in a stochastic approach, local sensitivity analysis was performed for the best model out of 20 random realizations (see Rogiers *et al.*, 2013c; Chapter 8). As the small source/sink (the rivers, canal and drains) sensitivities and fixed head boundary condition

sensitivities are addressed by Gedeon & Mallants (2012), we do not include them in the current analyses. Similar to the findings of these authors, the composite scaled sensitivities (Hill & Tiedeman 2007; Figure 9.1) identify the recharge (R_TOTAL) as the most sensitive parameter. R_TOTAL is multiplied by the initial recharge values (taken from the HYDRUS-1D (Šimůnek *et al.*, 2005) calculations of Leterme & Mallants 2012) for crop, meadow, coniferous and deciduous forest landcover types (R_CROP; R_MEADOW; R_CONIF; R_DECID), as well as the built area recharge R_BUILT which is expressed as a percentage of R_MEADOW. Following the recharge parameters, the horizontal hydraulic conductivity parameters are the most important (HK_HUF3D and HK_HUF1D). These are multiplication factors for the hydraulic conductivity fields respectively within the HUF3D and HUF1D parts of the model (see Rogiers *et al.*, 2013c; Chapter 8). The VANI_HUF3D parameter (vertical anisotropy for the HUF3D part of the model) is more sensitive than expected (Gedeon & Mallants 2012), probably due to the use of vertical head differences additional to the head observations for evaluating model performance. The VANI_HUF1D parameter seems to be the least sensitive, which corresponds to the extremely low lower aquifer VANI parameters' sensitivity of these authors. The conductance parameters (COND_RIVER; COND_CANAL; COND_DRAIN; COND_HOOI) form the least important parameter group.

Given these results, the fact that we do not want to modify the highly-data conditioned HUF3D marginal distribution by changing HK_HUF3D or VANI_HUF3D, and the fact that we are trying to quantify the effects of subsurface heterogeneity, we decide to keep the following parameters for the uncertainty analysis: HK_HUF1D, R_TOTAL, and R_BUILT. The ratios between the crop, meadow, coniferous and deciduous forest recharge will thus remain constant (and equal to those obtained by Leterme & Mallants (2012)), as well as the VANI_HUF1D and conductance parameters.

The reason to keep the HUF3D parameters fixed and optimize recharge is that we assume the marginal distributions of these parameters to be correct., and to find the exact balance between recharge and *K*, we need age tracer data, which is currently not available.

Table 9.1 summarizes all global model parameters, and indicates which ones are considered in the stochastic approach. We will consider 3 global parameters in the updated model (HK_HUF1D, R_TOTAL, R_BUILT), and two cases for the reference model: 6 and 11 global parameters. The model parameters that correspond to the VANI_HUF1D and conductance parameters, are thus kept constant in the uncertainty analysis.

The 11 parameters that remain for the reference model are a single recharge parameter (recharge is thus homogeneous), eight hydraulic conductivity and two vertical anisotropy parameters. For compatibility with Gedeon & Mallants (2012) and Gedeon *et al.* (2011), we consider additionally the use of 6 parameters for the reference model, in which there is one K_h value for the upper aquifer (the units within the upper aquifer are scaled to that value), one for the Kasterlee Clay aquitard, and one for the lower aquifer (the units within it are again scaled relatively to that value).

The considered parameter ranges are provided in Table 9.1 as well. For *K* parameters, we consider all values within one order of magnitude from the reference model parameterization. For the VANI parameters, isotropy is considered as the lower bound (value of one), and the upper bound is based on

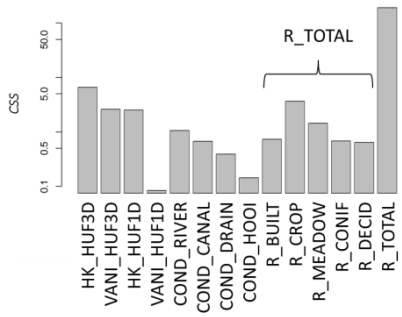

Figure 9.1: Composite scaled sensitivities (CSS) for all model parameters. HK: horizontal *K*; COND: conductance; R: recharge; HUF3D, HUF1D: sections of the flow model (see Chapter 8).

Table 9.1: Overview of the global model parameters and parameter ranges.

Reference model with 11 parameters		Reference model with 6 parameters		Updated model	
Parameter	**Range**	**Parameter**	**Range**	**Parameter**	**Range**
HK_Molup	[0.1,10] x 8.29 m/d	**HK_upaq**	[0.1,10] x 13.824 m/d	*HK_HUF3D**	*1*
HK_Molcl	[0.1,10] x 0.14 m/d				
HK_Mollw	[0.1,10] x 13.82 m/d				
HK_KastCl	[0.1,10] x 0.25 m/d	**HK_KastCl**	[0.1,10] x 0.25056 m/d		
HK_DistCl	[0.1,10] x 0.73 m/d	**HK_lowaq**	[0.1,10] x 5.184 m/d	**HK_HUF1D**	[1,5]
HK_Diest	[0.1,10] x 5.18 m/d				
HK_Dessel	[0.1,10] x 0.52 m/d				
HK_Berchem	[0.1,10] x 0.016 m/d				
VANI_upaq	[1,2]	**VANI_upaq**	[1,2]	*VANI_HUF3D**	*1*
VANI_KC	[1,5000]	**VANI_KC**	[1,5000]		
*VANI_Diest**	*3*	*VANI_Diest**	*3*	*VANI_HUF1D**	*1.62*
*VANI_Deep**	*1*	*VANI_Deep**	*1*		
*COND_RIVER**	*10*	*COND_RIVER**	*10*	*COND_RIVER**	*10*
*COND_CANAL**	*1*	*COND_CANAL**	*1*	*COND_CANAL**	*1*
*COND_DRAIN**	*10*	*COND_DRAIN**	*10*	*COND_DRAIN**	*10*
*COND_HOOI**	*1*	*COND_HOOI**	*1*	*COND_HOOI**	*1*
RECHARGE	[237,496] mm	**RECHARGE**	[0.5,2]*307 mm	**R_BUILT**	[0.4,1]
				*R_CROP**	*496 mm*
				*R_MEADOW**	*307 mm*
				*R_CONIF**	*237 mm*
				*R_DECID**	*376 mm*
				R_TOTAL	[0.6,1]

** Not considered stochastically*

previous experience with the model and incorporation of CPT data (Rogiers *et al.*, 2012b). The recharge parameter bounds are based on the meadow recharge, or the total range of recharge for the different land cover types, as estimated by Leterme & Mallants (2012). The recharge multipliers in the updated model use the values obtained by these authors as an upper bound. The treatment of the spatially distributed parameters within the updated model is outlined below.

9.2.2. Uncertainty quantification of the flow model

9.2.2.1. Adaptive Metropolis

The choice of an effective proposal distribution (both its scale and orientation) for the random walk Metropolis algorithm (Metropolis *et al.*, 1953) is essential for obtaining reasonably accurate results in a limited amount of time (Haario *et al.*, 1999). Adaptive McMC algorithms (*e.g.* Haario *et al.*, 2001; Ter Braak and Vrugt, 2008) use the history of the McMC sampling process for tuning online the proposal distribution, which results in large gain of efficiency compared to non-adaptive approaches. In this work, we employ the adaptive Metropolis (AM) algorithm of Haario *et al.* (2001) for sampling the global parameters. In the AM algorithm, the proposal distribution is a Gaussian distribution centred on the current position of the Markov chain, and with a covariance matrix C_t at iteration t calculated from all previous positions in the chain $X_0, \ldots, X_{t-1}$. As the final algorithm for sampling the posterior should not be adaptive, the use of all previous positions in the chain ensures that C_t will converge after a certain number of iterations and that the adaptation will vanish. This however requires an initial period with initial covariance C_0 and length $t_0 > 0$ so that

$$C_t = \begin{cases} C_0, & t \leq t_0 \\ s_d cov(X_0, \dots, X_{t-1}) + s_d \varepsilon I_d, & t > t_0 \end{cases} \quad [9.1]$$

with s_d a scaling parameter, and ε a very small constant and I_d the d-dimensional identity matrix to ensure C_t will not become singular. For the initial covariance matrix we use a diagonal matrix with variances obtained by a coefficient of variation of 5%, and start with $s_d = (2.38)^2/d$, which has been shown to be an optimal value in case of Gaussian proposals and targets (Gelman *et al.*, 1996). We do include an initial period of length t_f in which we adapt the scaling factor for tuning the acceptance rate *AR* by

$$s_{d,t} = \begin{cases} s_{d,t-1} \times 0.9, & AR < AR_{min} \\ s_{d,t-1}, & AR_{min} \leq AR \leq AR_{max} \\ s_{d,t-1} \times 1.1, & AR > AR_{max} \end{cases} \quad [9.2]$$

with AR_{min} = 0.1 and AR_{max} = 0.5 in our case, where *AR* is calculated from the last 10 samples. As all parameters in the considered model versions have a physical meaning, we would like to prevent them from moving outside of their uniform prior range. Therefore we use the reflection strategy for handling the parameter bounds, provided in Table 9.1, in which the excursion of the parameter outside of the range is added (substracted) from the exceeded lower (upper) bound.

The candidate point Z is accepted with probability (Metropolis *et al.* 1953)

$$\alpha(Z, X_{t-1}) = \min\left(1, \frac{P(Z|D)}{P(X_{t-1}|D)} \frac{Q(X_{t-1}|Z)}{Q(Z|X_{t-1})}\right) \quad [9.3]$$

Where D signifies the measurement data, $P(Z|D)/P(X_{t-1}|D)$ is the posterior density ratio between the proposed sample Z and the previous sample X_{t-1} and $Q(X_{t-1}|Z)/Q(Z|X_{t-1})$ is the ratio of the backward transition probability of creating X_{t-1} from Z to the forward transition probability of proposing Z from X_{t-1}. The latter ratio is equal to 1 in AM as the Gaussian proposal distribution is symmetric. If Z is accepted, then X_t is set to Z; otherwise the chain remains at the same location and $X_t = X_{t-1}$.

Assuming uniform priors for the model parameters, together with Normally distributed, homoscedastic and uncorrelated residuals for our two types of measurement data the (unnormalized) posterior density $P(X|D)$ can be summarized as

$$P(X|D) \propto \left([\sum_{k=1}^{n}(h_k - h'_k)^2]^{-\frac{n}{2}}\right) \times \left([\sum_{k=1}^{n}(g_k - g'_k)^2]^{-\frac{m}{2}}\right) \quad [9.4]$$

with $h_1, \dots, h_n$ the head observations and $g_1, \dots, g_m$ the head differences. Using Equation 9.4 implies the assumption that the variance of the error residuals is a good approximation of the measurement error of the observation data.

9.2.2.2. Adaptive spatial resampling

Perturbing spatially distributed parameter fields, obtained using geostatistical simulation, for sampling in the framework of McMC or optimization methods is not straightforward as the prior spatial model should be honoured. The spatially distributed parameters in the HUF3D part of the flow model should be perturbed in such a way that the non-stationary multivariate geostatistical model by Rogiers *et al.* (2013c; Chapter 8) is not violated. A possible solution to this problem was proposed by Mariethoz *et al.* (2010) who used a general transition kernel, called Iterative Spatial Resampling (ISR), that preserves any spatial model, making use of the conditional simulation capabilities of most existing geostatistical algorithms.

The approach consists of the following steps: 1) Generate an initial realization $X_0=X_{t-1}$ and evaluate the model likelihood (in this case using Eq. 9.4), 2) sample randomly a fraction φ of X_{t-1}, 3) generate a proposal realization Z by conditional simulation using the same geostatistical model, and conditioning on samples from X_{t-1}, 4) evaluate the model likelihood, calculate the acceptance probability α (Eq. 9.3), 5) if the model is accepted set X_t = Z, otherwise set $X_t = X_{t-1}$, and 6) return to step 2).

The only parameter that can be tweaked in this approach is φ. Mariethoz *et al.* (2010) use φ=0.1 or even smaller values for optimization.

The optimal φ for obtaining a proper acceptance rate in McMC sampling is of course highly problem-specific. If many informative observations of the modelled system are available, the problem is well constrained, and the optimal φ is likely to be higher. On the other hand, if only few observations are available, a smaller φ will allow the necessary exploration of the parameter space. After running some initial tests, we clearly needed much higher φ values than reported before for application of ISR. Therefore, we added an initial period where φ is

adapted to tune the acceptance rate *AR*, similar to the adaptation of the scaling parameter s_d in the AM algorithm (Eq. 9.3), but with the addition or substraction of 2% instead of the relative increase of 10% of the parameter value

$$\varphi_t = \begin{cases} \varphi_{t-1} + 0.02, & AR < AR_{min} \\ \varphi_{t-1}, & AR_{min} \leq AR \leq AR_{max} \\ \varphi_{t-1} - 0.02, & AR > AR_{max} \end{cases} \quad [9.5]$$

We started at the upper limit φ = 0.99 to ensure creating variability from the very beginning of the chain, and got 0.85 after an adaptive period of 200 iterations.

Jeong *et al.* (2012) modified the ISR algorithm by increasing the efficiency of sampling X_{t-1}. The probability of sampling the individual points of X_{t-1} is modified according to the spatially distributed model error. In a groundwater flow model, it makes sense to increase the probability of changing *K* values near locations with a high misfit between predicted and observed heads. As the model error can only be evaluated at the observation locations, we interpolate this error with inverse distance weighting with power 0.1 and vertical anisotropy of 200. The spatially distributed sampling probability Ψ is then calculated as

$$\Psi = \frac{exp(-r^2)}{\sum exp(-r^2)} \quad [9.6]$$

with r the spatially distributed residual. Comparable to the AM covariance adaptation, we use the average residual of X_0, …, X_{t-1} to make the adaptivity vanishing.

For evaluating the posterior density, Eq. 9.4 is used.

9.2.2.3. Block sampling

As the AM algorithm is suited for the global model parameters, and the ASR algorithm will be used for the spatially distributed parameter realization, and generating a proposal for the latter requires many times the CPU time for generating a global model parameter proposal, it would be efficient if we could avoid rejecting the geostatistical realization on the basis of a bad global model parameter proposal. Therefore, we apply a block sampling strategy in which the global parameters and the geostatistical field are updated each one at the time. This is accomplished in a simple way by considering the global model parameters on odd iterations, while keeping the geostatistical realization constant, and by performing ASR on all even iterations without changing the global parameters.

9.2.2.4. McMC sampling vs optimization mode

As the forward model takes between 6 and 18 minutes, and the geostatistical realization, which is performed every other iteration due to the block sampling approach, takes about 7 minutes with the φ values used in this paper, the number of iterations that can be performed is limited. Nevertheless, the results will be more representative for the posterior as those obtained by performing an equal amount of iterations using a pure Monte Carlo approach.

In order to assess convergence of the Markov chains, we use the convergence diagnostic of Geweke (1992). This convergence diagnostic for Markov chains is based on a test for equality of the means of the first and last part of a Markov chain. The first 10% and the last 50% are commonly used for this. If the samples are drawn from the stationary distribution of the chain, the two means are equal and Geweke's statistic has an asymptotically standard normal distribution. The test statistic is a standard Z-score for which we provide the p-values.

As convergence of the McMC chain is very computationally expensive, certainly with the updated model, we use an additional optimization approach to explore the best performing part of the parameter space. The algorithm can easily be switched in optimization mode by only accepting better performing models when evaluating the likelihood ratio (Eq. 9.3). We make a second change to run the algorithm in optimization mode by setting the *AR* bounds to [0.1,0.9] in Eqs 9.2 and 9.5. To explore the best performing part of the parameter space and retain some variability as well, we did a series of optimization runs starting from the same initial model, with a fixed amount of 300 iterations each.

The final algorithm resulting from the combination of the above approaches is summarized in Figure 9.A.

9.2.3. The transport model

For the transport simulations we use the MT3DMS code (Zheng & Wang, 1999) together with the LMT6 package (Zheng *et al.*, 2001)

that translates the MODFLOW flow output to the required input for MT3DMS. For solving the dispersion term, the implicit Finite-Difference method using the General Conjugate-Gradient (GCG) solver is used. The advection term solution is discussed below.

9.2.3.1. Transport parameters

As we only consider stochasticity of the solute transport due to the flow field, we only create one deterministic porosity and dispersivity field for the transport simulations.
As we will be doing steady-state transport, the porosity has no impact on the transport results. For the transient simulations performed for investigating the effect of numerical dispersion, the porosity is taken from the reference model parameterization for transport simulations. The porosity used is the average total porosity measured on a large set of 100 cm³ core samples from the eight boreholes discussed by Beerten *et al.* (2010), including the seven boreholes studied by Rogiers *et al.* (2012a; Chapter 6; 2013b; Chapter 5), as no data is available concerning effective porosity. As we are again interested in results at a steady-state, the importance of the porosity will be minor for these transient simulations as well. The different values are shown in Table 9.2.

Physical dispersion was not accounted for in the transport calculations by Gedeon *et al.* (2011). Instead the numerical dispersion, arising through the used advection solution scheme, was used as a surrogate for the physical dispersion. The underlying assumption is that if numerical dispersion is smaller than the true physical dispersion, estimates of a maximum concentration will be on the conservative side. Depending on the scale considered for estimating the physical dispersion, this approach can be criticized, especially in the numerical models with 50x50 m² horizontal discretization of the numerical grid.
Therefore, we try to obtain a set of realistic dispersivity estimates based on the outcrop investigations by Rogiers *et al.* (2013a; Chapter 3) and literature data, displayed in Figure 9.2A. As the horizontal flow component is more important than the vertical one in our groundwater flow model, we only consider dispersivities for horizontal flow. We project the longitudinal dispersivities (α_L) estimated for the outcrops to the 250 m-scale, using the linear relationship derived from literature data (Figure 9.2)

$$log_{10}(\alpha_L) = -1.1727 + 0.8471\ log_{10}(scale) \quad [9.7]$$

-Start with initial global parameters and geostatistical realization
-Run the forward model
-Calculate the likelihood
-Start main loop with t=1
A If t is odd, go to B, else go to C
B Use the AM algorithm for the global parameters
-Update C_t and s_d if necessary (Eqs 9.1 & 9.2)
-Generate a proposal distribution and draw a sample for the global X_t parameters
-Check the prior parameter ranges and use reflection if the bounds are violated
-Go to D
C Use the ASR algorithm for the geostatistical realization
-Update φ and Ψ if necessary (Eqs 9.5 & 9.6)
-Sample a fraction φ from the distributed X_{t-1} parameters
-Generate a new geostatistical realization conditioned on the primary and secondary data, as well as the sampled X_{t-1} parameters
-Go to D
D Evaluate the performance of X_t
-Run the forward model
-Calculate the likelihood ratio (Eqs 9.3 & 9.4)
-If X_t is rejected: replace X_t by X_{t-1}
-t=t+1
-Go to A

Figure 9.A: Summary of the final algorithm.

Table 9.2: Average total porosity for the different hydrogeological units

Hydrogeological unit	Average total porosity
Quaternary & Mol Upper Sands	0.37
Mol Clayey Layer	0.40
Mol Lower & Kasterlee Sands	0.41
Kasterlee Clay	0.40
Diest Clayey Top	0.42
Diest Sands	0.42
Dessel Sands	0.40
Berchem & Voort Sands	0.39

Note that there is a discrepancy between 2D and 3D dispersivity, and no correction was performed. This projection to larger scales however has a much bigger influence on the obtained values. We use the 250 m-scale because the heterogeneity at 250 m and beyond is represented explicitely in the model through the geostatistical realization (only one in five cells is used in the horizontal direction). This local-scale dispersion that we include in the transport model should therefore only represent the sub-250 m-scale heterogeneity, if we consider solute plume extents beyond 250 m.

For the reference model, the use of macroscopic dispersion is however required. Therefore we repeat the above approach for the 1.5 km-scale, which is the scale of the solute plume. Both 250 m-scale dispersivity and 1.5 km-scale dispersivity are tested with the reference model flow fields.

Gelhar (1993) provides different approaches to estimate macrodispersion parameters based on spatial heterogeneity of flow and K, as well as local dispersivity in function of K. Within a non-stationary anisotropic medium, application of these approaches is however not straightforward. Because of the importance of logarithmic K variance, we decide to use the empirical relationship with α_L and the logarithmic K variance divided by the scale of the outcrop investigation, the outcrop width (Figure 9.2B), for estimating both local-scale and macroscopic dispersivity parameters within the transport model. We use the CPT-based logarithmic K variance for this purpose, and perform simple kriging inter-/extrapolation to the entire numerical grid. The scale representative for the CPT-based logarithmic K variance is adjusted in order to obtain the same geometric mean K value as for the outcrops. The result for the first numerical layer is presented in Figure 9.3A, and clearly illustrates the presence of the homogeneous Mol Sands in the northeastern corner of the study area. Figure 9.3B provides the histograms of the dispersivities estimated from the CPT data, and that resulting from interpolation to the numerical grid. Note that the bulk of α_L values are between $10^{0.8}$ and $10^{1.3}$ m, and that the very low and high values from the CPT dataset only occur in a very limited part of the model, or disappear by the smoothing effect of the interpolation.

The ratios of horizontal transverse dispersivity α_{TH} and vertical transverse dispersivity α_{TV} to α_L generally are in the range 0.1-0.01 and 0.01-0.001 respectively; the average values determined from the outcrops amount 0.06 and 0.006 (Rogiers *et al.*, 2013a; Chapter 3). We can however not use these ratios in the current model as the scale of heterogeneity explicitely respresented by the geostatistical realization is one meter instead of 250 m in the horizontal direction. Therefore we rescale these ratios using the ratio between 250 m-scale α_L and one meter-scale α_L as estimated from Eq. *9.*7. The resulting values are respectively $5.6\ 10^{-4}$ and $5.6\ 10^{-5}$. For the 1.5 km-scale dispersivity, the average outcrop values are used.

9.2.3.2. Initial and boundary conditions

For the initial conditions, the concentration is set to zero in the entire model. The disposal facility design includes two buildings, which will be covered by an eartch cover at the final stage of the disposal

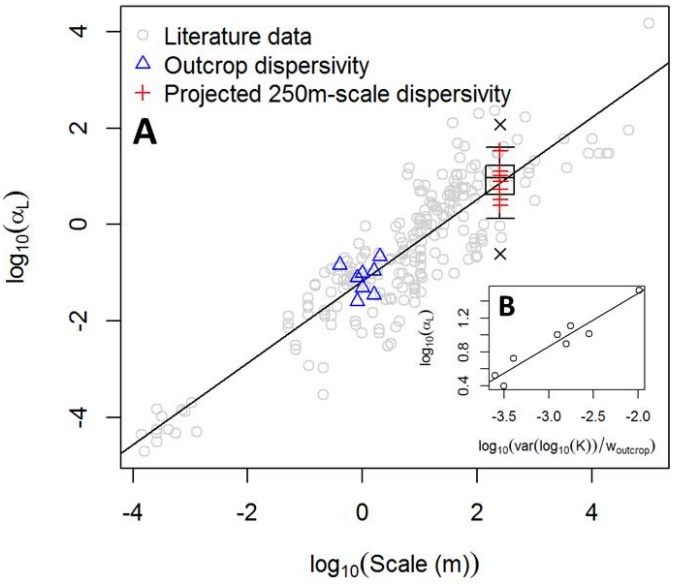

Figure 9.2: A) Boxplot of the 250 m-scale dispersivity estimates from CPT-based K variance, and literature data for longitudinal dispersivity (Baumann *et al.*, 2002, 2005, 2010; Gelhar *et al.*, 1985, 1992; Kim *et al.*, 2002; Mallants *et al.*, 2000; Schulze-Makuch *et al.*, 2005). The outcrop dispersivities derived by Rogiers *et al.* (2013a; Chapter 3) and their corresponding projections to the 250 m scale are displayed as well. B) The linear relationship between logarithmic geometric K variability, scaled by the outcrop width, and logarithmic 250 m-scale longitudinal dispersivity, based on the outcrop data of Rogiers *et al.* (2013a; Chapter 3).

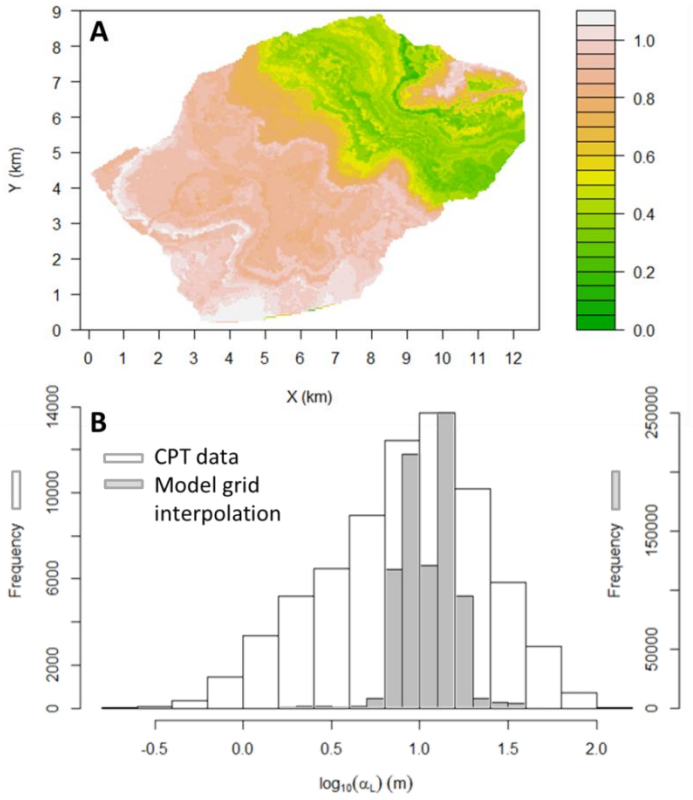

Figure 9.3: A) 2D α_L field for the first numerical layer, and B) histograms of the α_L estimates based on the CPT dataset and the interpolation to the numerical grid.

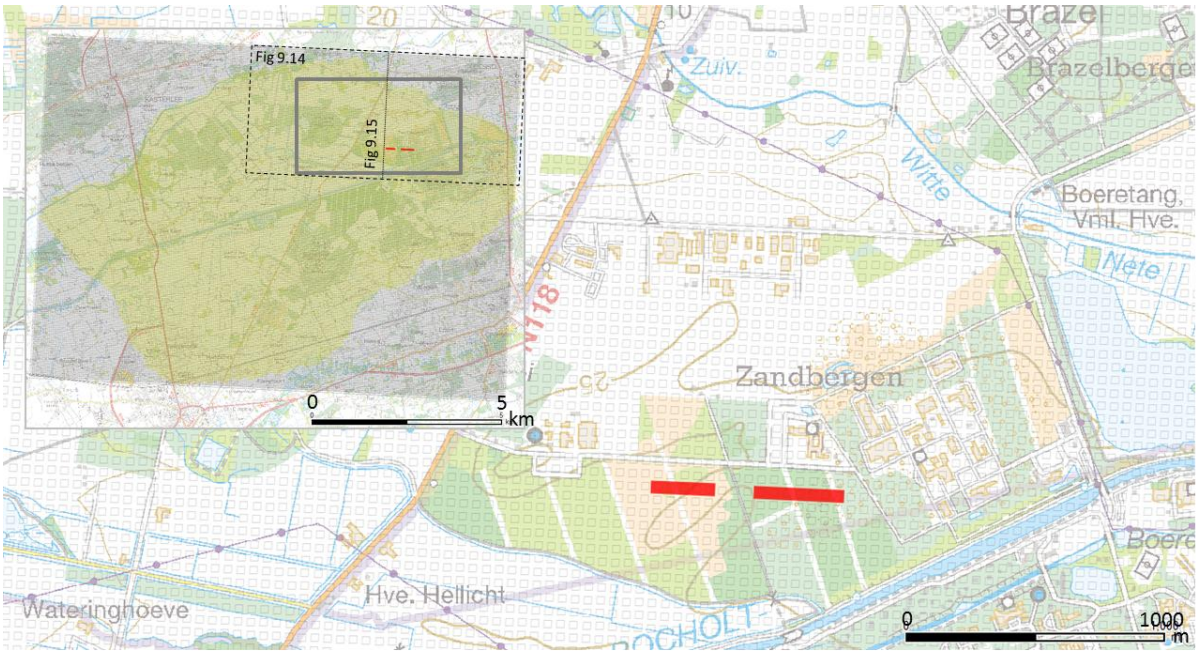

Figure 9.4: Location of the disposal tumuli within the groundwater flow and solute tranport model, and the topographical map displaying the surroundings of the tumuli.

project. The location of the resulting disposal tumuli is indicated on Figure 9.4. The solute source is defined as a constant mass loading flux, which equals 99932 Bq/y in the 5 grid cells corresponding to the western tumulus in Figure 9.4, and 71380 Bq/y in the 7 grid cells corresponding to the eastern tumulus. This mass loading corresponds to a total input of 10^6 Bq/y in the model, which is divided equally among the two tumuli.

Given the slight differences in the saturated thickness of the source cells, both within and between the two tumuli, as well as the flow through the cells, slight differences in the steady-state concentration that is reached might occur.

The no-flow boundaries in the groundwater flow model are translated to zero mass flux boundaries in the transport model. The fixed and general head boundaries, as well as the rivers and drains are translated into specified mass flux boundaries, with mass flux QC with Q determined by MODFLOW, and C the respective cell concentration. The dispersive flux across boundaries is always assumed to be negligible.

9.2.3.3. Advection solution and numerical dispersion

The Eulerian Finite-Difference method has been used for the reference model transport simulations. A steady-state solution is obtained by solving only the spatial derivatives in the partial differential equation, which allows to obtain results in at most a few minutes. Therefore, this steady-state advection solution is very appealing for performing the transport calculations for the obtained flow model ensembles. However, to minimize numerical dispersion with this Finite-Difference advection scheme, fine spatial discretization may be necessary to satisfy the Peclet number constraint

$$Pe = \frac{v_x \Delta x}{D_{xx}} = \frac{\Delta x}{\alpha_L} \leq 2 - 4 \qquad [9.8]$$

with v_x the effective velocity, Δx the grid cell size, and D_{xx} the dispersion coefficient. All Peclet numbers > 1 in fact represent the case where numerical dispersion exceeds the magnitude of physical dispersion, and using the steady-state finite difference advection solution would not lead to conservative safety assessment in terms of maximum concentrations. With a plume extent of ~1500 m, we obtain a dispersivity of ~33 m, which leads to Pe = 1.5, indicating that numerical dispersion would indeed exceed

physical dispersion, but the magnitude is however similar. For the detailed-scale transport simulation by Gedeon *et al.* (2011), with 10x10 m² horizontal discretization, *Pe* = 0.3, and the result would indeed be conservative in terms of maximum concentrations. This reasoning is however only valid for the reference model with homogeneous hydrogeological layers. For the updated model, the considered scale for dispersivity is 250 m, which results in α_L values, mainly between 6.3 and 20 m (Figure 9.3B), or *Pe* = 2.5 - 7.9. Therefore, we should look for other advection solution schemes that are less prone to numerical dispersion.

As the computational time required for obtaining a steady-state with the Eulerian-Lagrangian method of characteristics (MOC) transient solution scheme is relatively high, we however decide to keep using the steady-state finite difference solution for looking at the model ensemble output. If we acknowledge the effect of numerical dispersion for interpreting the results of simulating advection and advection with physical dispersion, we cannot make conclusions on safety assessment, but we can compare the different models as *Pe* stays constant. For more accurate quantification of the solute migration, we select a few representative model realizations, and use the MOC advection solution scheme, with the output obtained after simulating 100 years of solute plume migration.

It is stressed here that we never claim to model reality with the steady-state finite-difference solution. The detailed-scale transport simulation by Gedeon *et al.* (2011) delivers a result in terms of maximum concentrations that are strongly affected by numerical dispersion (this latter is largely influenced by the mesh size of the model grid); not reality. There are however indications that this numerical dispersion is smaller than the physical dispersion which can be expected in reality. The simulations with the current updated model provide a feasible approach to obtain indicative steady-state transport results from large sets of models, which can be compared, but solute spreading is consistently overestimated, and does not represent reality.

9.2.3.4. Solute plume characteristics

To image the solute plume properties, and the differences between the different model ensembles, we use a series of plume characteristics and projections: a total activity map and profile, the plume volume, the dilution rate, which is defined below, and the maximum concentration in function of the distance from the disposal facility:

For looking at the spatial distribution of the solute plume, we use a map projection of the total activity in the entire Neogene aquifer, as well as a profile projection (~north-south), along the rows of the model. Each pixel in these total activity projections of the plume represents the sum of activities in all of the cells in that particular cell sequence, and hence the total activity is expressed per m² (cell area). The activity is calculated by multiplying the concentration with the cell volume and total porosity, as specified before. To capture the volume of the solute plume in a single numerical value, we look at the volume of the aquifer in which the concentration exceeds an arbitrary value of 0.01 Bq/m³.

As the near field simulations consider thousands of years, the solute fluxes from the disposal facility are considered to spread instantaneous in the geosphere (or at least the upper aquifer) at each time step of the near field model. This instantaneous spread of the solute within the geosphere allows to use a steady-state transport solution. The outcome is evaluated by the dilution rate (Figure 9.5). The dilution rate is defined as the total source flux to the geosphere (we use 10^6 Bq/y) divided by the maximum average concentration within the upper aquifer, at least 70 m from the disposal facility. It is the main safety indicator for the geosphere component in the safety assessment within the category A project. The reason for averaging the concentration in the upper aquifer is that a potential well would always draw water from the entire upper aquifer. This dilution rate is then used for converting the solute fluxes, outputted at each time step from the near field model, to a concentration that serves as input for the biosphere modelling, to assess the actual dose rates on man and environment.

Note that the term dilution, as used in this case, is not the specific result of pore-scale dispersion, as it is commonly used. It is just an indicator on the solute transport outcome. We in fact consider spreading by using dispersion arising from hydraulic conductivity heterogeneity.

The maximum concentration within a single cell in function of the perpendicular distance from the center of the disposal facility is more

informative than the dilution factor, but does not account for mixing due to pumping from a hypothetical well. As other pathways to the biosphere might be important as well, we do include this maximum concentration profile with distance.
All these characteristics of the simulated solute plume should be interpreted with caution, because of the numerical dispersion that influences the results considerably. Relative differences between the model ensembles should however provide an idea of the different processes and the effect of heterogeneity. The MOC solutions provide more meaningful absolute values.

9.3. Results and discussion

9.3.1. Flow model ensemble results

For evaluating the flow model ensemble results, we use the log-likelihood evolution in function of the number of forward model simulations, and the posterior parameter distributions after thinning of the chain results by taking each 20th sample to reduce the autocorrelation of the results. For the optimization mode, only one sample is retained from each individual chain, and thinning is not required.
Table 9.3 summarizes the computations that have been performed. An AM McMC chain was used for the reference model with 6 and 11 parameters, and the updated model parameter space was explored using the combination of AM and ASR by block sampling, both in McMC and optimization mode. An additional non-adaptive ISR chain was used to verify the properties of the ISR algorithm and the used starting φ of 0.99. The reference model chains are started from the reference parameterization; the updated model chains from the best solution of some initial trial runs of the McMC mode sampling, to start closer to the posterior.
Figure 9.6 shows the results for the reference model with 6 parameters. According to the log-likelihood evolution of the chain, the deterministic reference model parameterization, which is used as starting point, seems to be part of the posterior. Most posterior parameter distributions are centered around the starting point. Only the lower aquifer K_h (HK_lowaq) seems to have some bias between the reference value and its marginal posterior distribution.
The evolution of the 6 global parameters is shown in Figure 9.7. Visually the chain seems to have converged after the first 1000 iterations. The p-values for the Geweke (1992) diagnostic after removal of the first 1000 iterations are given in Table 9.4. These indicate that VANI_KC might not have converged yet, and the continuation of the sampling is recommended for better approximation of the posterior. Given the already large amount of forward model runs, we use the current results as an approximation of the posterior.
Figure 9.8 shows the results for the reference model with 11 parameters. From the log-likelihood evolution it becomes clear that this is a more difficult inverse problem, and convergence of the chain might not be reached. The initial model again seems to be part of the posterior, but the range of log-likelihoods seems not to have stabilized yet. All marginal posterior parameter distributions indicate that the initial values are highly probable.
The evolution of the 11 global parameters is shown in Figure 9.9. Visually the chain seems not to have converged after the first 5000 iterations. The p-values for the Geweke (1992) diagnostic are given in Table 9.5, and indicate that the mean is not stable for several parameters. Similar to the reference model chain with 6 parameters, we use these results as an approximation of the posterior.
The results from the updated model McMC sampling are shown in Figure 9.10. The log-likelihood however seems to decrease almost systematically from the initial optimal value (Figure 9.10A). The algorithm does not seem to be able to stay in the better performing part of the parameter space. To investigate the reason for this, we started an additional chain with constant global parameters, and without adaptation, *i.e.* an ISR chain, with φ = 0.99. As this chain seems to be able to recover from the initial drop in log-likelihood, we suspect that the adaptation of φ (Eq. 9.5) is too large in the initial algorithm for the current problem. After the adaptive period of the first 200 iterations, φ indeed dropped to 0.85 in the initial chain, and that in just a few iterations. This might induce too large proposal jumps, but a few more thousands of samples would be needed before drawing any conclusions. A solution could be to temper the adaptation of φ or to put a lower bound on the parameter.

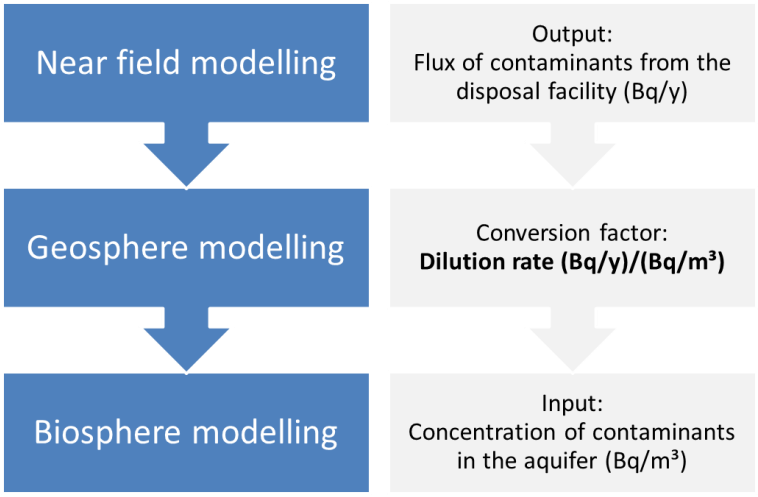

Figure 9.5: Framework in which the dilution rate is used, and which leads to the division of a flux by a concentration.

Table 9.3: Overview of sampling calculations for the different stochastic flow models.

Model	Sampling	Global parameters	Spatially distributed parameters	Forward model evaluations
Reference model	AM	6	-	7281
Reference model	AM	11	-	8544
Updated model	Block sampling: AM + ASR (McMC mode)	3	41183 x 2 (K_h & VANI)	693
Updated model	ISR	-	41183 x 2 (K_h & VANI)	767
Updated model	Block sampling: AM + ASR (optimization mode)	3	41183 x 2 (K_h & VANI)	300 x 8

Table 9.4: p-values for the Geweke (1992) convergence diagnostic for the reference model with 6 parameters, after removal of the first 1000 iterations.

Parameter	p value
RECHARGE	2.86E-04
HK_upaq	1.15E-05
HK_KastCl	3.90E-02
HK_lowaq	4.52E-02
VANI_upaq	4.89E-04
VANI_KC	9.41E-01

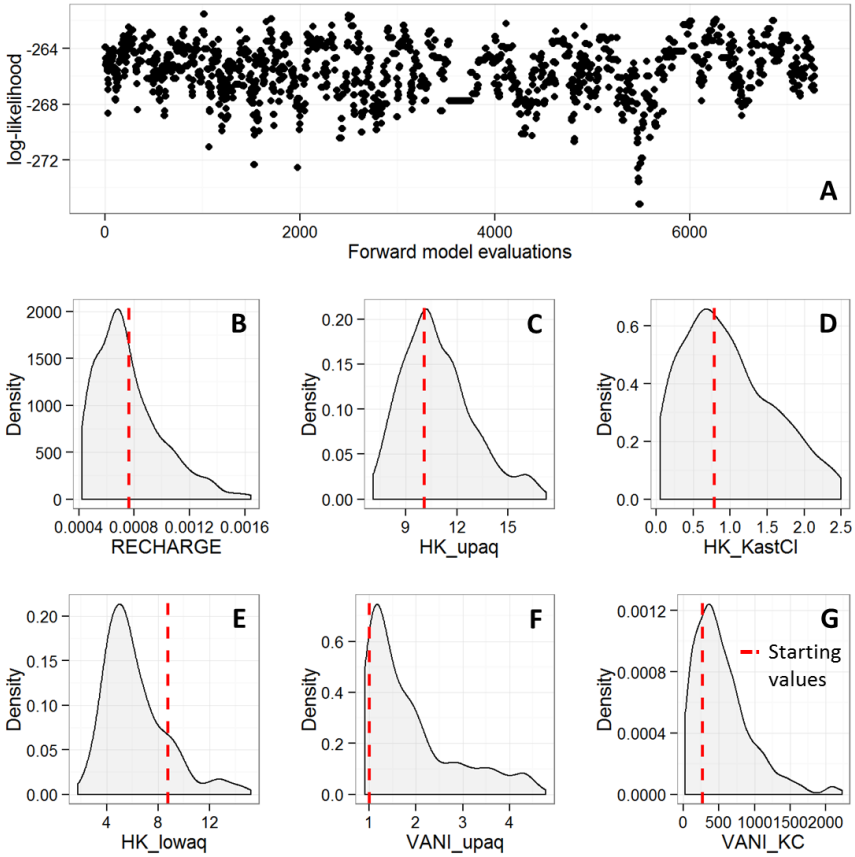

Figure 9.6: McMC sampling results from the reference model with 6 parameters: A) log-likelihood in function of the number of forward model evaluations, B-G) posterior marginal parameter distributions.

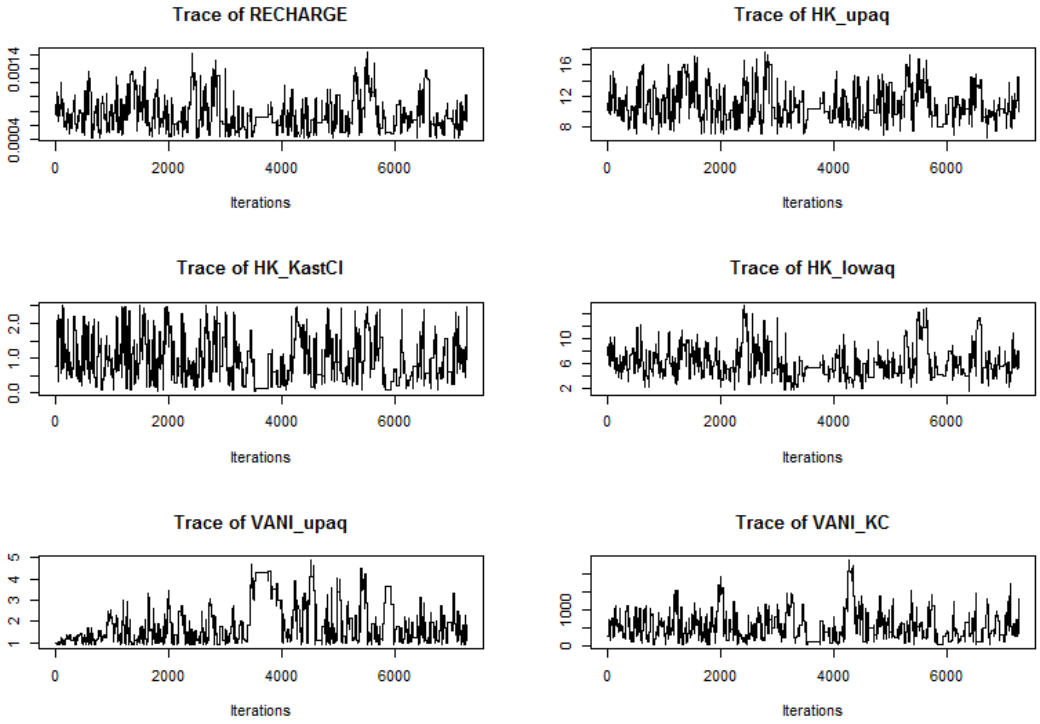

Figure 9.7: Evolution of the 6 global parameters within the reference model with the number forward model runs (iterations).

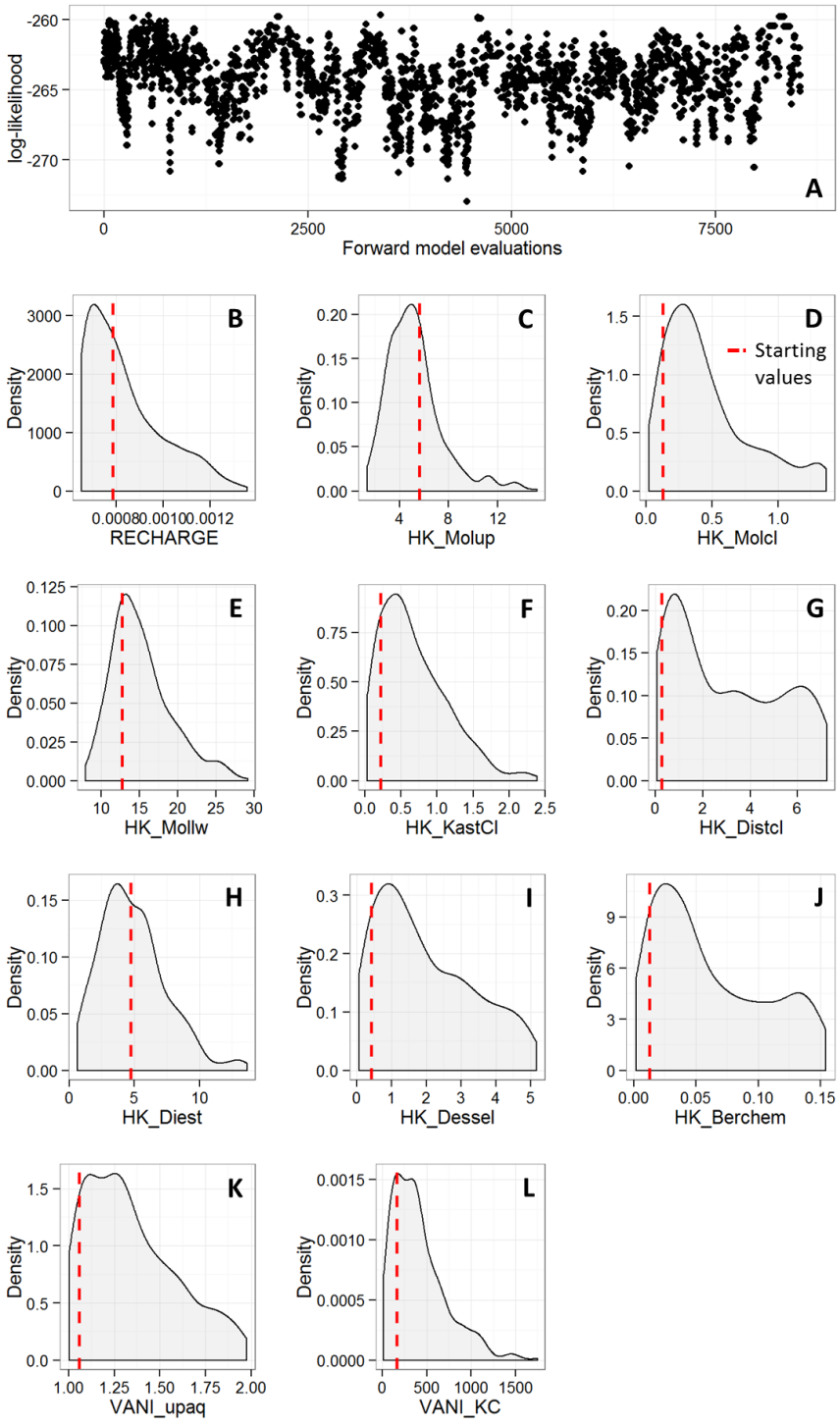

Figure 9.8: McMC sampling results from the reference model with 11 parameters: A) log-likelihood in function of the number of forward model evaluations, B-L) posterior marginal parameter distributions.

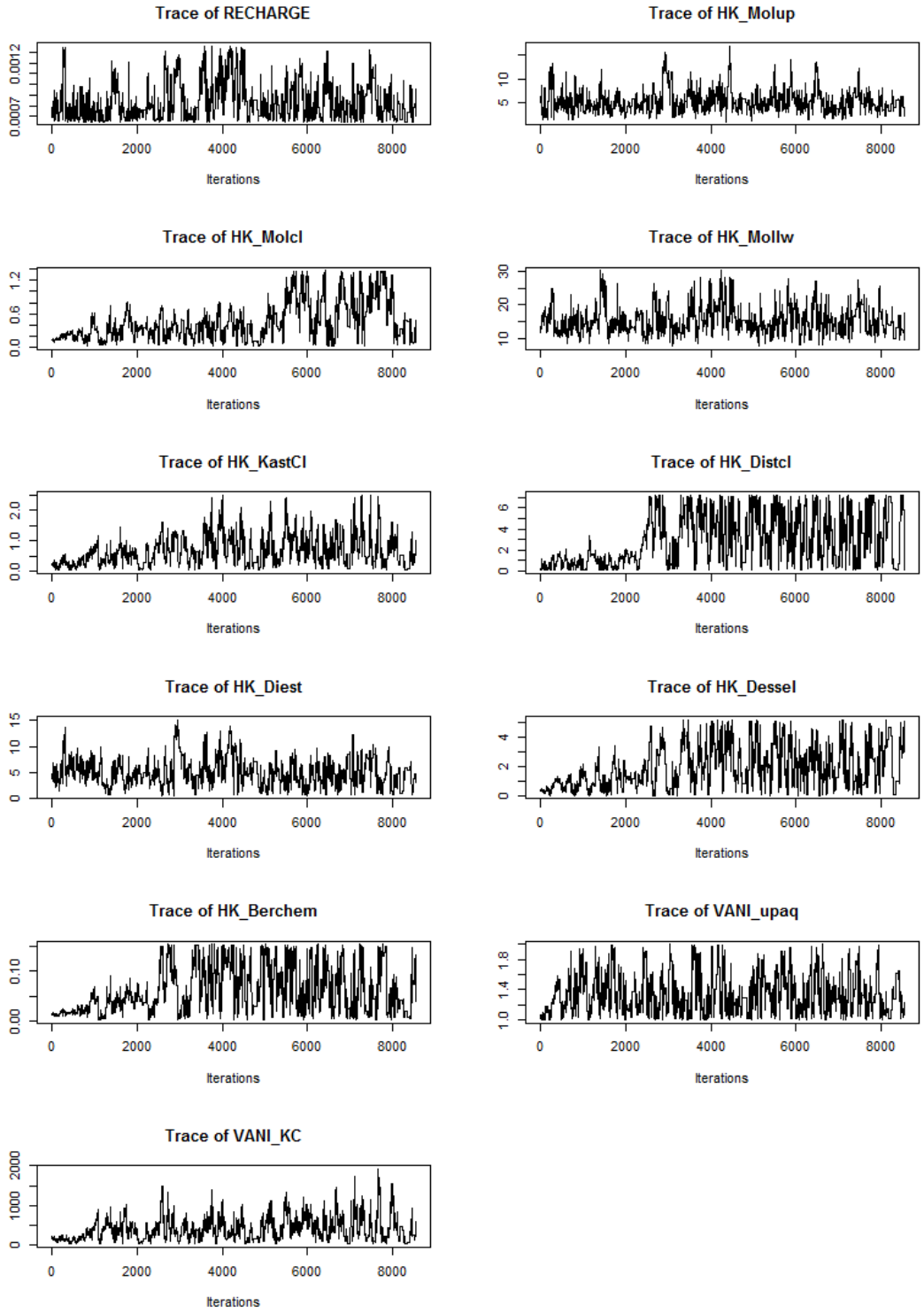

Figure 9.9: Evolution of the 11 global parameters within the reference model with the number of forward model runs (iterations).

The posterior marginal parameter distributions indicate that the overall recharge is probably not below 80% of the estimates of Leterme & Mallants (2012), but the observation data is not very informative on the recharge in urban areas (R_BUILT). The multiplier for K_h in the HUF1D part of the model (HK_HUF1D) indicates that some scale discrepancy between the upscaled data-based estimates and the effective model parameters still exists, and amounts to a factor 2 to 4.

The result of eight optimization mode chains, starting from the same optimal distribution, are displayed in Figure 9.11. The chains all start from the same model, and new proposals are only accepted if the log-likelihood increases. After 300 forward model evaluations, the chains are interrupted, and the final model is kept. The posterior marginal global parameter distributions are very different from the previous results, but given the low number of samples, no conclusions can be drawn.

The means of the *K* fields of three example hydrogeological unit flow layers are provided in Figure 9.12, for 50 samples of the McMC chain, and the 8 optimization mode samples. The difference between the two sets of samples seems to be mainly slightly larger extent of the very low K_v values within the Kasterlee Clay aquitard. Compared to the random realizations from Chapter 8, the *K* values seem to be higher, mainly in the SW of the study area, which might indicate that important regional differences in inter-aquifer connectivity exist.

A comparison between the deterministic reference model parameterization, and the best performing model from the optimization mode sampling is provided in Figure 9.13. Some of the outliers found by Rogiers *et al.* (2013c; Chapter 8) have improved, but two large head residuals remain. The dimensionless scaled sensitivities of the river and drain conductances (not shown) for these particular observations indicate however that further improvement might be obtained when including these parameters.

Table 9.5: p-values for the Geweke (1992) convergence diagnostic for the reference model with 11 parameters, after removal of the first 1000 iterations.

Parameter	p-value
RECHARGE	1.33E-01
HK_Molup	2.59E-01
HK_Molcl	1.49E-11
HK_Mollw	1.02E-01
HK_KastCl	2.57E-03
HK_Distcl	1.10E-23
HK_Diest	6.48E-01
HK_Dessel	2.37E-10
HK_Berchem	6.71E-11
VANI_upaq	7.13E-02
VANI_KC	7.98E-01

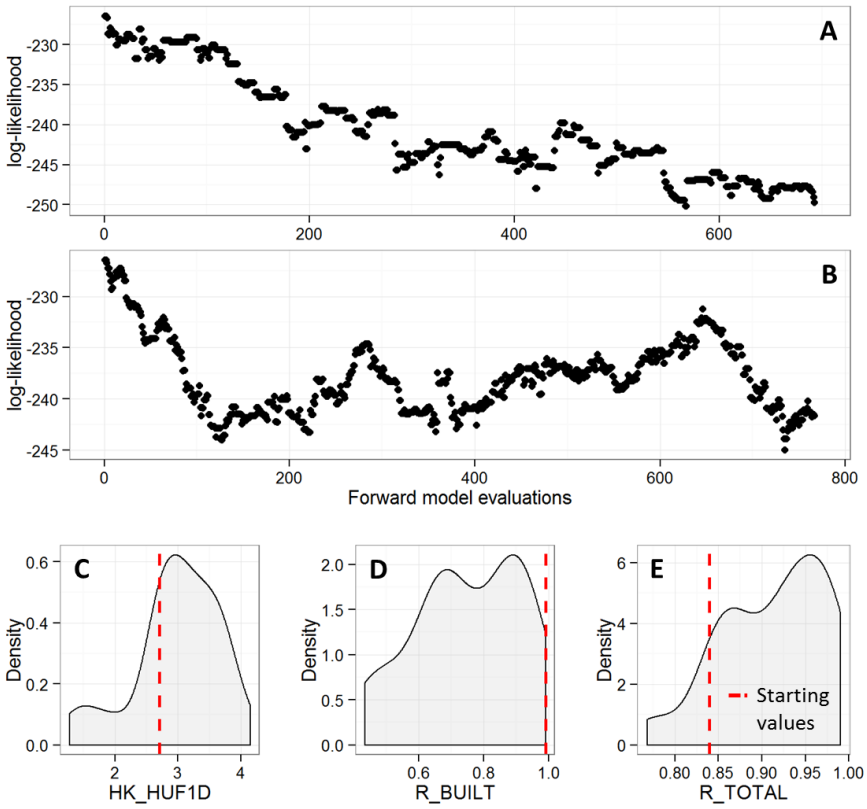

Figure 9.10: McMC sampling results from the updated model with 3 global parameters: A) log-likelihood in function of the number of forward model evaluations, B) ISR chain without adaptivity to check the effect of a constant φ, C-E) posterior marginal global parameter distributions (from A, as the global parameters were constant in B).

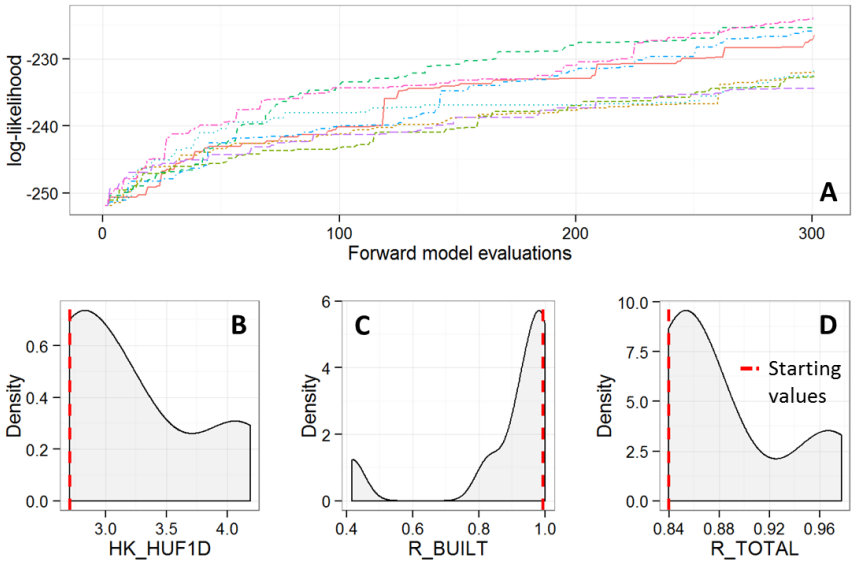

Figure 9.11: Optimization results for the updated model with 3 global parameters: A) log-likelihood in function of the number of forward model evaluations for each chain, B-D) posterior marginal parameter distributions.

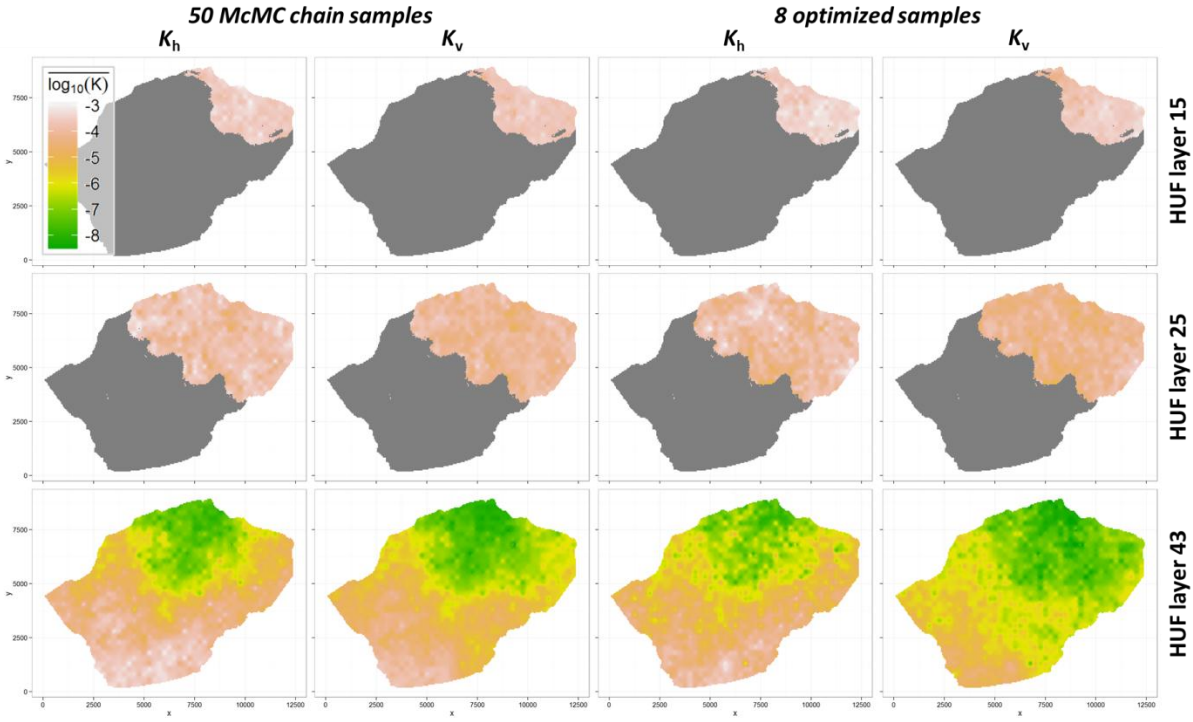

Figure 9.12: Mean K_h and K_v fields for 3 hydrogeologic unit flow (HUF) layers from 50 McMC chain samples and 8 samples of the optimization mode sampling. HUF layer 15 represents the high *K* coarse Mol Upper sands; HUF layer 25 is located in the center of the upper aquifer, and HUF layer 43 is located within the Kasterlee Clay aquitard.

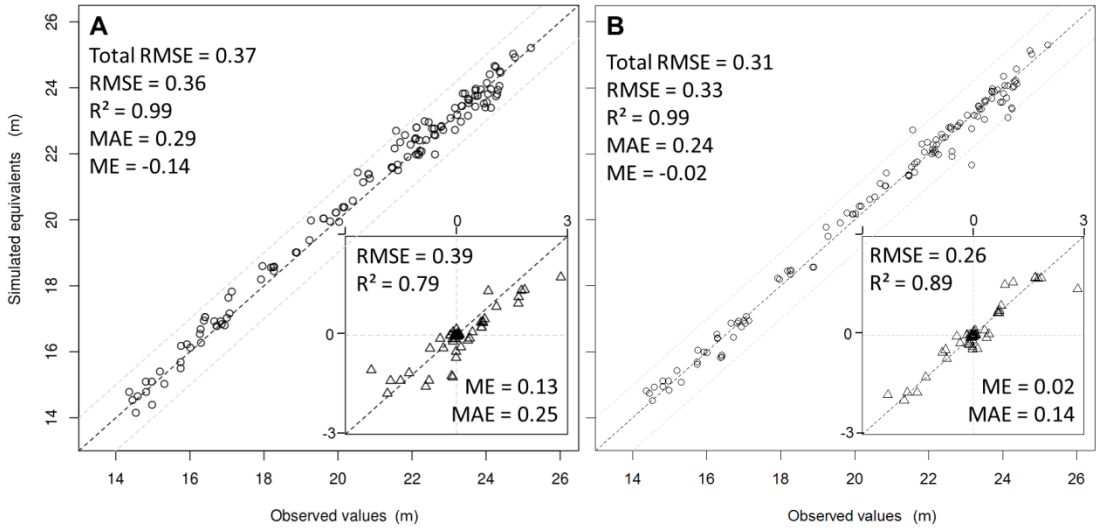

Figure 9.13: Predicted versus observed heads for A) the deterministic reference case parameterization with homogeneous hydrogeological layers, and B) the maximum likelihood model from the optimization mode sampling. The dashed lines represent a residual of 1 m. The inset shows the vertical head differences used in the objective function. RMSE: root mean squared error; MAE: mean absolute error; ME: mean error.

9.3.2. Transport model ensemble results

Samples should be drawn from the flow model ensembles in order to represent the ensemble by a reasonable number of realizations, that is still easily processed by the transport model. Therefore we chose to select 50 flow fields from the end of the McMC chains, as well as the eight realizations obtained from the optimization mode calculations. Moreover, the flow solutions of 20 random geostatistical realizations are used as well, to quantify the effect of inverse conditioning on the head observations. This leads to a total of 178 flow solutions used for transport calculations with the steady-state finite-difference advection scheme.

9.3.2.1. Total activity risk maps and profiles

The total activity risk maps are displayed in Figure 9.14, the profiles in Figure 9.15. The probability of exceeding the threshold seems to be the lowest in the reference model results, both in the map and profile views (Figure 9.14A,B, Figure 9.15A,B). The solute plume migrates downward in the upper aquifer, and searches its way to the nearest rivers. The main solute mass stays in the upper aquifer, but some manages to penetrate the aquitard, reaches the lower aquifer, and flows back up again near the rivers. The insertion of heterogeneity in the model creates much larger uncertainties concerning the spatial location of the solute plume (Figure 9.14C, Figure 9.15C), which are again reduced by inverse conditioning (Figure 9.14D, Figure 9.15D). The limited set of optimized samples (Figure 9.14E, Figure 9.15E) seems to provide similar results than the updated model McMC samples.
The steady-state mass of solute from the eastern tumulus seems to be negligible to that of the western tumulus. This might be caused by the shorter flow path to the nearest model sinks, as well as the presence of the high *K*, homogeneous coarse Upper Mol sands in the eastern corner of the study area. Moreover, the mass loading flux per unit area is also lower for this tumulus.

9.3.2.2. Maximum concentration in function of distance

The maximum concentration in function of the perpendicular distance to the center of the tumuli is shown in Figure 9.16. The reference model ensembles seem to provide the results centered around the deterministic reference model output (Figure 9.16A,B). The parameterization with 6 parameters results in a larger uncertainty than that of the one with 11 parameters. The 20 random realizations show an even larger uncertainty than the reference model with 6 parameters (Figure 9.16C), even though the global parameters are constant in this case. This indicates the importance of the spatially heterogeneous parameter field. The samples from the McMC chain of the updated model show a rather narrow distribution, with concentrations systematically below that of the reference model parameterization (Figure 9.16D). The eight optimized models show very little variability in maximum concentrations, with even lower values than the McMC samples (Figure 9.16E). This indicates that accounting for heterogeneity through conditional geostatistical simulation, as well as inverse conditioning on the observations reduces the maximum concentration observed within the Neogene aquifer. The numerical dispersion is expected to affect all these results in a similar way, so that the relative differences between the different model ensembles and reference model parameterization are likely to be valid.

9.3.2.3. Solute plume volume

The numerical dispersion makes the absolute values of the derived plume volumes not very reliable. The relative differences are however interesting. The posterior solute plume volume distributions, as quantified by considering the cells with concentration above 0.01 Bq/m³, are shown in Figure 9.17.
The reference model distributions seem te be centered around the deterministic reference model parameterization (Figure 9.17A,B). Due to the heterogeneity introduced by the 20 random geostatistical realizations with constant global model parameters, the spread of the solutes within the Neogene aquifer is considerably higher (Figure 9.17C). The McMC chain samples from the updated model confirm these larger plume volumes, but the range of values is not as high as that of the random realizations (Figure 9.17D), probably due to the inverse conditioning. The eight optimized models of course show a narrow distribution, which does not contain the deterministic

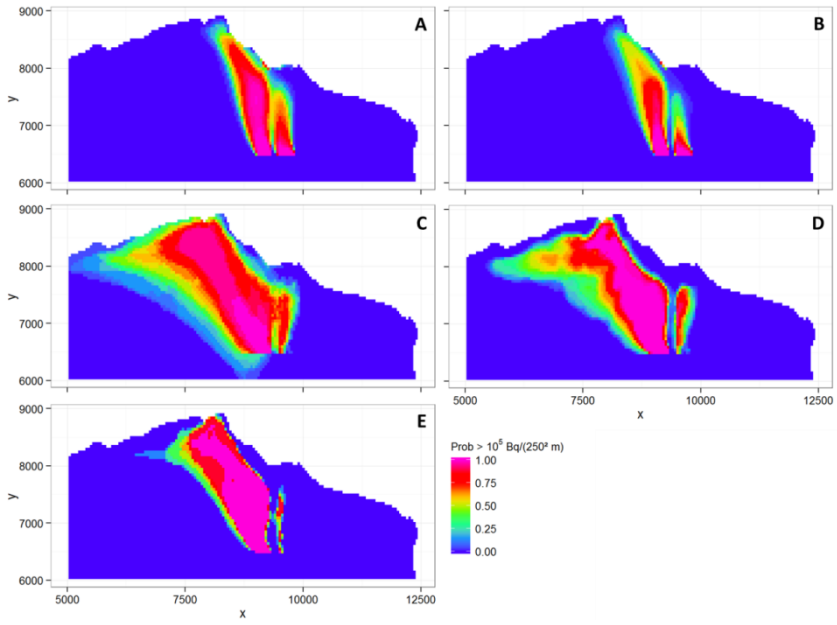

Figure 9.14: Risk maps indicating the probability for the total activity (summed for all layers) within the entire Neogene aquifer to exceed 10^5 Bq/250 m², for the steady-state finite-difference advection scheme. A) 50 samples from the reference model with 6 parameters, B) 50 samples from the reference model with 11 parameters, C) 20 random realizations, D) 50 samples from the updated model McMC chain, E) 8 optimized results from 8 optimization mode chains. Considerable numerical dispersion has influenced these results. The exact location of these maps is displayed in Figure 9.4.

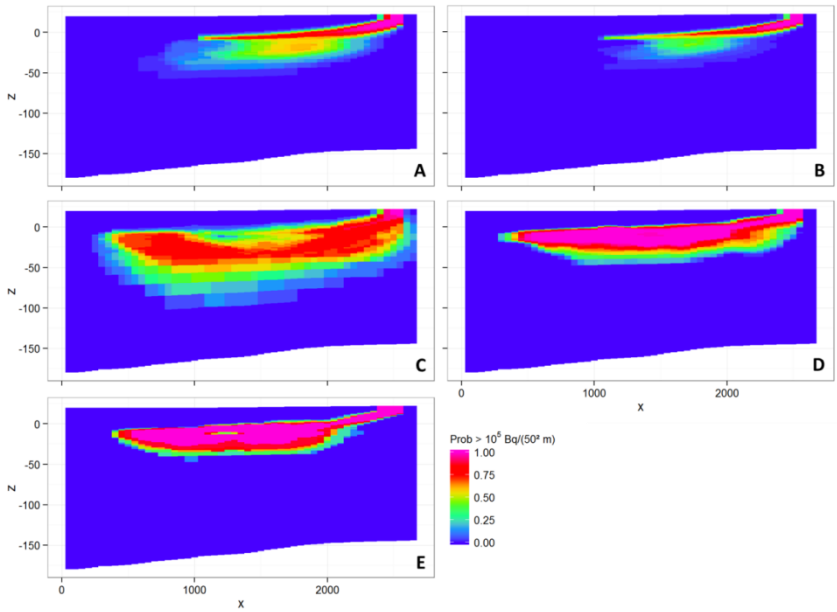

Figure 9.15: Risk profiles indicating the probability for the total activity (summed for all columns) within the entire Neogene aquifer to exceed 10^5 Bq/50 m², for the steady-state finite-difference advection scheme. A) 50 samples from the reference model with 6 parameters, B) 50 samples from the reference model with 11 parameters, C) 20 random realizations, D) 50 samples from the updated model McMC chain, E) 8 optimized results from 8 optimization mode chains. Considerable numerical dispersion has influenced these results. The orientation of these profiles is displayed in Figure 9.4.

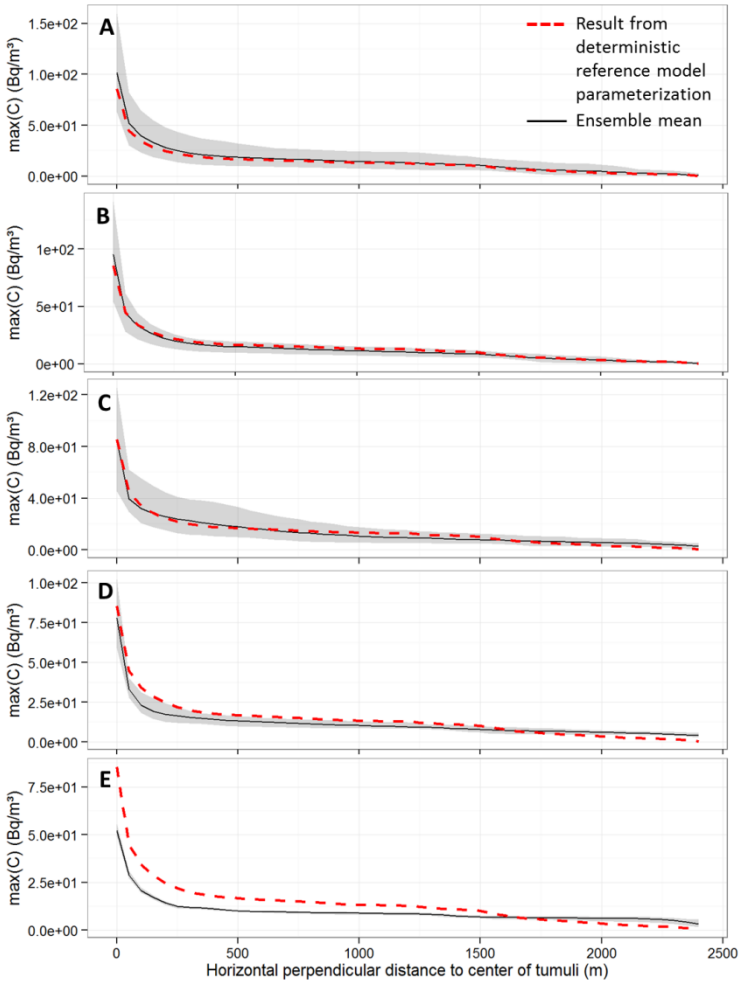

Figure 9.16: Posterior distributions of the maximum concentration within the Neogene aquifer, in function of the horizontal perpendicular distance from the center of the tumuli. A) 50 samples from the reference model with 6 parameters, B) 50 samples from the reference model with 11 parameters, C) 20 random realizations, D) 50 samples from the updated model McMC chain, E) 8 optimized results from 8 optimization mode chains.

reference model estimate (Figure 9.17E). Despite the effect of numerical dispersion, it is clear that the heterogeneity present in the subsurface causes a much larger portion of the Neogene aquifer to be affected than what is estimated with the reference models, or the deterministic reference model parameterization.

9.3.2.4. Dilution rate

The results for the dilution rate are displayed in Figure 9.18. The reference model and updated model McMC posterior distributions are all centered at the deterministic reference model parameterization result (Figure 9.18A,B,D). This indicates that for estimating the average concentration within the upper aquifer, close to the solute source, the errors introduced by considering homogeneous hydrogeological units are probably small. Random geostatistical realizations again lead to a high spread of the results, with dilution rates that are a lot lower than the deterministic reference model parameterization (Figure 9.18C). Inverse conditioning seems to rule out these low dilution rates, and the set of eight optimized models even is systematically above the deterministic reference model result (Figure 9.18E). These dilution rates are of course a product of the numerical dispersion, in combination with the heterogeneity that is inserted in the model. Therefore, the absolute values are not reliable,

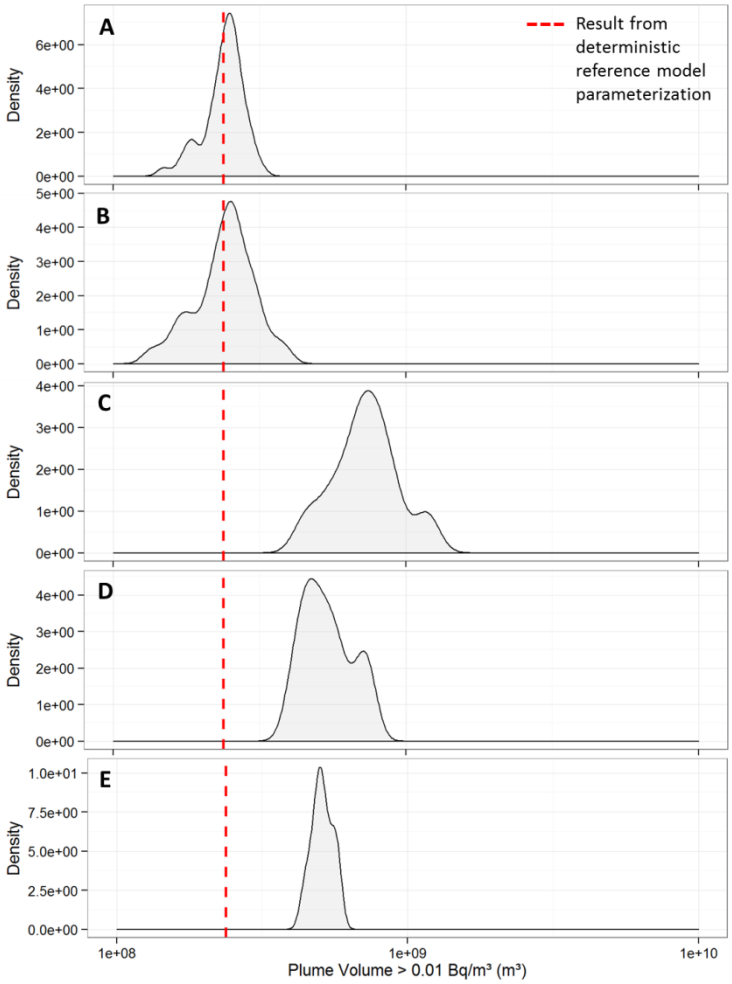

Figure 9.17: Posterior distribution of the contaminant plume volume, as quantified by a cutoff concentration of 0.01 Bq/m³. A) 50 samples from the reference model with 6 parameters, B) 50 samples from the reference model with 11 parameters, C) 20 random realizations, D) 50 samples from the updated model McMC chain, E) 8 optimized results from 8 optimization mode chains.

Table 9.6: Comparison of the dilution rate and plume volume results between the finite difference steady-state solution and the MOC results for a time of 100 years, using no physical dispersion.

Parameter	Advection + numerical dispersion			
	Reference model		Best performing realization	
	FD (steady-state)	MOC (100 years)	FD (steady-state)	MOC (100 years)
Dilution rate (Bq/y)/(Bq/m³)	6.18E+04	6.21E+04	1.04E+05	6.81E+04
Plume volume (m³)	2.37E+08	3.90E+07	4.99E+08	6.48E+07

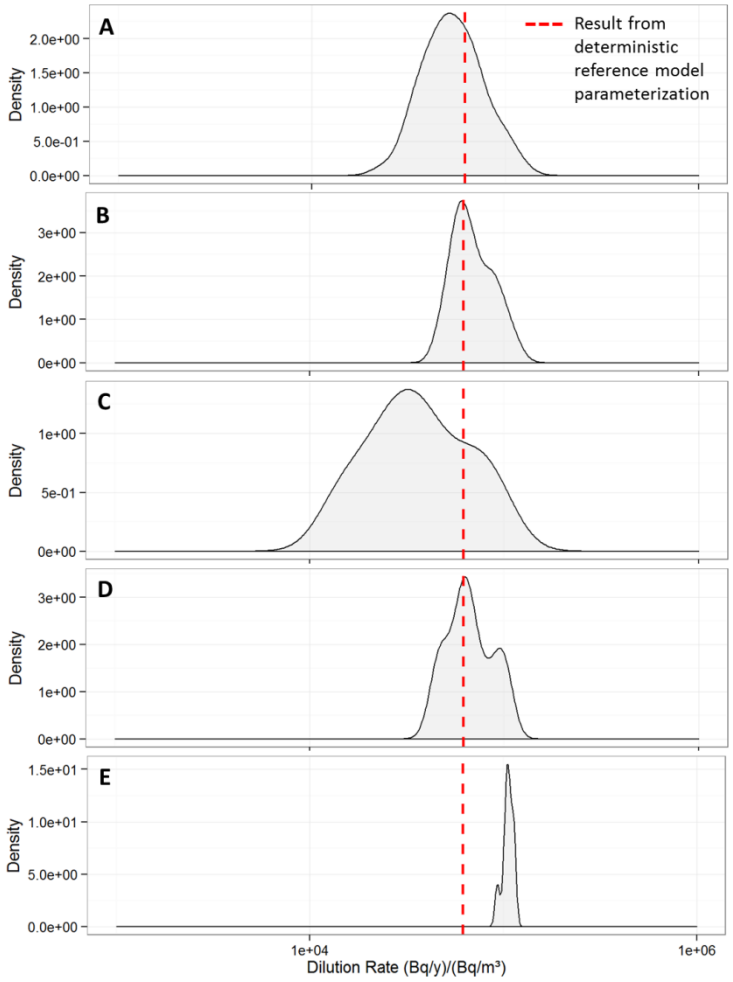

Figure 9.18: Posterior distribution of the dilution rate in the upper aquifer. A) 50 samples from the reference model with 6 parameters, B) 50 samples from the reference model with 11 parameters, C) 20 random realizations, D) 50 samples from the updated model McMC chain, E) 8 optimized results from 8 optimization mode chains.

Table 9.7: Comparison of the dilution rate and plume volume results from the MOC simulations for a time of 100 years, using local-scale or macroscopic dispersion.

Parameter	Advection + numerical dispersion + 250m-scale physical dispersion		Advection + numerical dispersion + 1500 m-scale physical dispersion
	Reference model	Best performing realization	Reference model
	MOC (100 years)		
Dilution rate (Bq/y)/(Bq/m³)	5.89E+04	6.54E+04	5.17E+04
Plume volume (m³)	4.80E+07	8.21E+07	8.70E+07

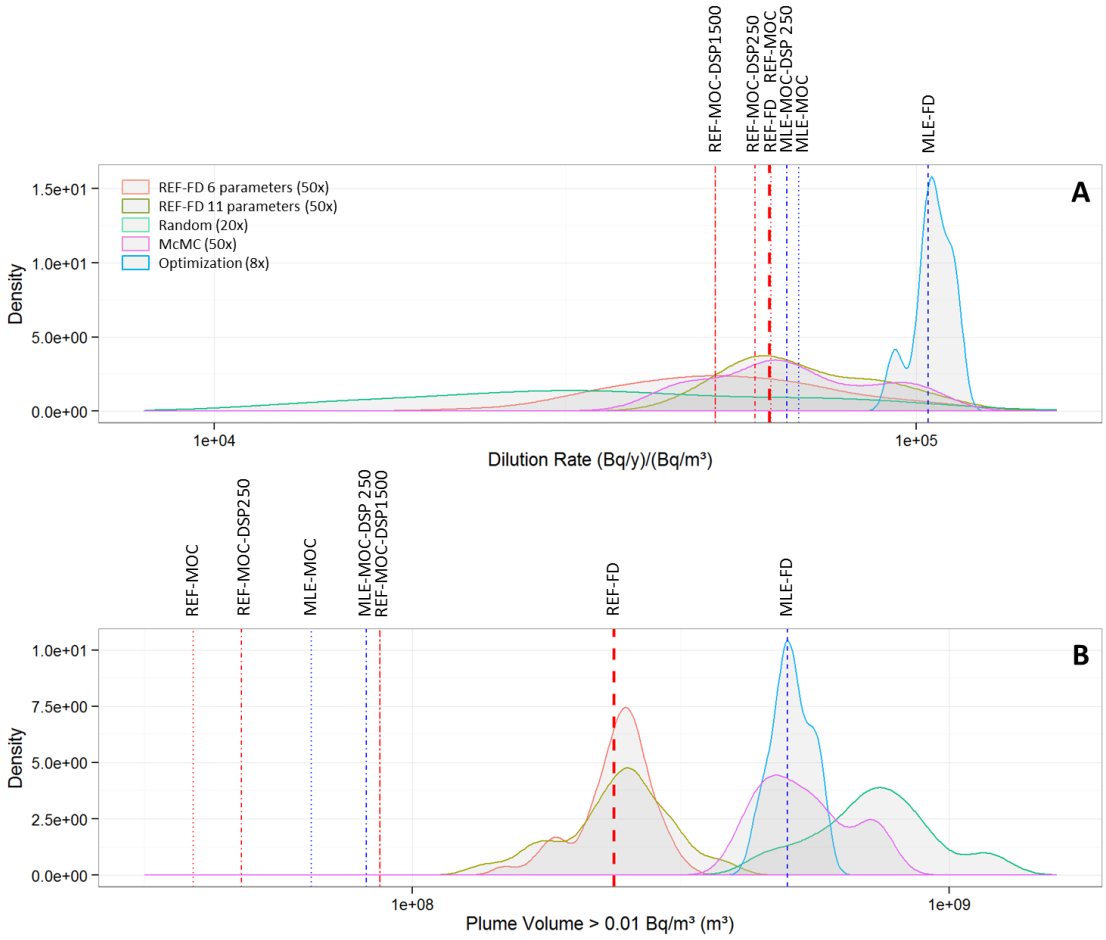

Figure 9.19: Overview of the dilution rates and plume volumes of the finite difference (FD) and method of characteristics (MOC) solutions for the deterministic reference model parameterization (REF) and the updated model optimization mode sample with maximum likelihood (MLE). In both cases, 250 m-local-scale dispersion was used, and for the reference model the 1.5 km-scale macroscopic dispersion was tested as well. The model ensemble results from the FD simulations are shown as reference.

and should be assessed using other advection solution schemes or finer numerical discretization. Nevertheless, the insertion of heterogeneity in the model seems to yield similar dilution rates than the reference model, and the best performing part of the parameter space seems to yield consistently higher values.

9.3.3. MOC simulation results

The method of characteristics was used for solving advection in transient simulations of 100 years to check the difference with the results of the finite difference solution. This was done both with and without the estimated physical dispersion discussed above. The results for the dilution rate and the volume of the solute plume are given in Table 9.6 for the simulations without physical dispersion, and in Table 9.7 for the simulations including physical dispersion. Figure 9.19 gives an overview of these results, as well as those from the flow model ensemble distributions.

For the reference model, the dilution rate seems to agree very well between the finite difference solution and the MOC result after 100 years. However, it was observed that the concentration decrease with increasing distance from the tumuli is highly irregular if no physical dispersion is specified. The inclusion of local-scale physical dispersion, estimated as outlined above, indeed lead to a more stable solution, with only a slightly lower dilution rate. The macroscopic dispersion simulation clearly provides a lower dilution rate. This can be explained by considerable increase of the solute

plume volume near the source area, which could result in a higher average concentration within the upper aquifer, but still a lower peak concentration, and is due to the very specific of the dilution rate. For the dilution rate, which is estimated close to the source, the 250 m-scale dispersion might however be more realistic. Given the heterogeneity within the updated model, the dilution rates estimated with this model are systematically higher than those of the reference model, which suggests that working with homogeneous hydrogeological units is conservative in terms of the dilution rate, if similar dispersion parameters are used.

For the plume volume, a large range of results is obtaiend, and the finite difference solution clearly makes an overestimation. The local-scale dispersion within the updated model, and the macroscopic dispersion within the reference model, which should more or less be equivalent, do provide a similar result.

9.4. Conclusions

At first sight, the combination of AM and ASR using block sampling seems not to be very effective for McMC sampling with the forward model used in this study. The results of using an ISR chain without adaptivity indicates that the adaptivity of the φ parameter, the fraction of constant nodes in the ISR/ASR method, might be too high, and additional tests should be performed to verify this. A lower bound or the use of smaller steps to keep the acceptance rate *AR* within the provided bounds might solve the problem. However, running the algorithm in optimization mode seems to work fine, and provides an alternate way for exploring the parameter space and the prediction uncertainty.

Despite the effect of numerical dispersion, the steady-state finite difference advection scheme solutions suggest that the reference model results in conservative estimates of the dilution rate, but the most pessimistic parameter set that is considered part of the posterior should be chosen to obtain a more robust probabilistic conservative estimate. The part of the aquifer that is affected by contamination, on the other hand, seems to be considerably underestimated by the use of homogeneous hydrogeological units.

The few transient simulations that make use of the method of characteristics to solve the advection seem to confirm the observations made from the steady-state finite difference solutions for the dilution factor, but the difference between the updated and reference model results is probably smaller than estimated from the model ensemble results. The inclusion of physical dispersion seems to suggest that the dilution rate in the upper aquifer might decrease with increasing dispersion. A large discrepancy exists between the finite difference and MOC estimates of the plume volume, but this might be due to the fact that simulating 100 years is probably far from reaching steady-state conditions. The local-scale dispersion in the updated model and the macroscopic dispersion within the reference model however provide similar plume volume estimates, as would be expected. Further analyses are required if highly accurate probabilistic results are required. Even with the use of Eulerian-Lagrangian methods, it is recommended to refine the numerical grid considerably to reduce numerical dispersion to a minimum. This can easily be done by selecting only a part of the flow model for transport calculations, as performed by Gedeon *et al.* (2011) for the detailed-scale modelling. Moreover, the geostatistical simulation approach by Rogiers *et al.* (2013c; Chapter 8) can easily be adjusted to create realizations at larger horizontal or vertical discretization, increasing the realism of small-scale variability within the model. Large-scale semi-parallel computing would however be required to achieve probabilistic results.

References

Beerten, K., Wemaere, I., Gedeon, M., Labat, S., Rogiers, B., Mallants, D., Salah, S., Leterme, B., 2010. Geological, hydrogeological and hydrological data for the Dessel disposal site. Project near surface disposal of category A waste at Dessel. STB-SIE(HYD) - Version 1, NIROND-TR 2009-05 E (p. 261).

Baumann, T., Müller, S., Niessner, R., 2002. Migration of dissolved heavy metal compounds and PCP in the presence of colloids through a heterogeneous calcareous gravel and a homogeneous quartz sand – pilot scale experiments. *Water Research* **36:** 1213–1223.

Baumann, T., Werth, C.J., 2005. Visualization of colloid transport through heterogeneous porous media using magnetic resonance imaging. *Colloids and Surfaces A: Physicochemical and Engineering Aspects* **265**: 2–10.

Baumann, T., Toops, L., Niessner, R., 2010. Colloid dispersion on the pore scale. *Water Research* **44**(4): 1246-54.

Beven, K., Freer, J., 2001. Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the GLUE methodology. *Journal of Hydrology* **249**(1-4): 11–29. doi:10.1016/S0022-1694(01)00421-8

Cui, T., Fox C., O'Sullivan M. J., 2011. Bayesian calibration of a large-scale geothermal reservoir model by a new adaptive delayed acceptance Metropolis Hastings algorithm, *Water Resour. Res.* **47**(10), W10521, doi:10.1029/2010WR010352.

Doherty, J., 2003. Ground water model calibration using pilot points and regularization. *Ground Water* **41**(2): 170–177.

Gedeon, M., Mallants, D., Vandersteen, K., Rogiers, B., Laloy, E., 2011. Hydrogeological modelling of the Dessel site - Overview report (p. 227).

Gedeon, M., Mallants, D., 2012. Sensitivity Analysis of a Combined Groundwater Flow and Solute Transport Model Using Local-Grid Refinement: A Case Study. *Mathematical Geosciences* **44**(7): 881–899. doi:10.1007/s11004-012-9416-3

Gelhar, L.W., Mantoglou, A., Welty, C., Rehfeldt, K.R., 1985. A review of field-scale physical solute transport processes in saturated and unsaturated porous media. Palo Alto, California: Electric Power Research Institute EPRI EA-4190 Project 2485–5.

Gelhar, L.W., Welty, C., Rehfeldt, K., 1992. A Critical Review of Data on Field-Scale Dispersion in Aquifers. *Water Resources Research* **28**(7): 1955-1974.

Gelhar, Lynn W., Stochastic Subsurface Hydrology, Prentice Hall, 390 pp., 1993.

Gelman, A.G., Roberts, G.O., Gilks, W.R., 1996. Efficient Metropolis jumping rules. In J.M. Bernardo, J.O., Berger, A.E., David, A.EM., Smith (eds), Bayesian Statistics V. pp. 599-608. Oxford: Oxford University Press.

Geweke, J., 1992. Evaluating the accuracy of sampling-based approaches to calculating posterior moments. 169-193. In Bayesian Statistics 4 (ed JM Bernado, JO Berger, AP Dawid and AFM Smith). Clarendon Press, Oxford, UK.

Haario, H., Saksman, E., Tamminen, J., 1999. Adaptive proposal distribution for random walk Metropolis algorithm. *Comput. Statist.* **14**: 375-395.

Haario, H., Saksman, E., Tamminen, J., 2001. An adaptive Metropolis algorithm. *Bernoulli* **7**(2): 223–242.

Haario, H., Laine, M., Mira, A., Saksman, E., 2006. DRAM: Efficient adaptive MCMC, *Stat. Comput.* **16**: 339–354.

Harbaugh, A. W., 2005. MODFLOW-2005, The US Geological Survey modular ground-water model—the Ground-Water Flow Process: US Geological Survey Techniques and Methods 6-A16, USGS, 2005.

He, X., Jiang, L., Moulton, J. D., 2013. A stochastic dimension reduction multiscale finite element method for groundwater flow problems in heterogeneous random porous media. *Journal of Hydrology* **478**: 77–88. doi:10.1016/j.jhydrol.2012.11.052

Hill, M.C., Tiedeman, C.R., 2007. Effective Groundwater Model Calibration: With Analysis of Data, Sensitivities, Predictions, and Uncertainty: Wiley and Sons, 464 p.

Jeong, C., Mukerji, T., Mariethoz, G., 2012. Adaptive Spatial Resampling Applied to Seismic Inverse Modeling. In Ninth International Geostatistics Congress (pp. 1–25).

Kim, D.-J., Kim, J.-S., Yun, S.-T., Lee, S.-H., 2002. Determination of longitudinal dispersivity in an unconfined sandy aquifer. *Hydrological Processes* **16**(10): 1955–1964.

Kourakos, G., Mantoglou, A., 2013. Development of a multi-objective optimization algorithm using surrogate models for coastal aquifer management. *Journal of Hydrology* **479**: 13–23. doi:10.1016/j.jhydrol.2012.10.050

Laloy, E., Vrugt, J. A., 2012. High-dimensional posterior exploration of hydrologic models using multiple-try DREAM(ZS) and high performance computing, *Water Resour. Res.* **48**(1), W01526, doi:10.1029/2011WR010608.

Laloy, E., Rogiers, B., Vrugt, J. a., Mallants, D., Jacques, D., 2013. Efficient posterior exploration of a high-dimensional groundwater model from two-stage Markov chain Monte Carlo simulation and polynomial chaos expansion. *Water Resources Research* **49**(5): 2664–2682. doi:10.1002/wrcr.20226

Leterme, B, Mallants, D., 2012. Simulation of evapotranspiration and groundwater recharge in the nete catchment accounting for different land cover types and for present and future climate conditions. Project near surface disposal of category A waste at Dessel. External report of the Belgian Nuclear Research Centre. SCK•CEN-ER-192.

Mallants, D., Espino, A., Hoorick, M., Feyen, J., Vandenberghe, N., Loy, W., 2000. Dispersivity estimates from a tracer experiment in a sandy aquifer. *Ground Water* **38**(2): 304–310.

Mariethoz, G., Renard, P., Caers, J., 2010. Bayesian inverse problem and optimization with iterative spatial resampling. *Water Resources Research* **46**(11), 1–17. doi:10.1029/2010WR009274

Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H., Teller, E., 1953. Equations of state calculations by fast computing machines. *J Chem. Phys.* **21**: 1087-1091.

Montanari, A., Shoemaker, C. A., Giesesen, N., 2009. Introduction to special section on Uncertainty Assessment in Surface and Subsurface Hydrology: An overview of issues and

challenges. *Water Resources Research* **45**(6): 2005–2008. doi:10.1029/2009WR008471

ONDRAF/NIRAS. 2010. Het cAt-project in Dessel. Een langetermijnoplossing voor het Belgische categorie A-afval. Retrieved from http://www.niras-cat.be/downloads/cAt_masterplan_NL_LOW.pdf on 07-12-2011.

Razavi, S., Tolson, B. A., Burn, D. H., 2012. Review of surrogate modeling in water resources. *Water Resources Research* **48**(7). doi:10.1029/2011WR011527

Refsgaard, J. C., van der Sluijs, J. P., Højberg, A. L., Vanrolleghem, P. A., 2007. Uncertainty in the environmental modelling process e A framework and guidance. *Water* **22**: 1543–1556. doi:10.1016/j.envsoft.2007.02.004

Rojas, R., Feyen, L., Dassargues, A., 2008. Combining the generalized likelihood uncertainty estimation (GLUE) and the Bayesian model averaging (BMA) to account for conceptual model uncertainty in groundwater modelling. *Water Resources Research* **44**(12): 1–16. doi:10.1029/2008WR006908

Robert, C.P., Casella, G., 2004. Monte Carlo Statistical Method, Springer.

Rogiers, B., Mallants, D., Batelaan, O., Gedeon, M., Huysmans, M., Dassargues, A., 2012a. Estimation of hydraulic conductivity and its uncertainty from grain-size data using GLUE and artificial neural networks. *Mathematical Geosciences* **44**(6): 739–763.

Rogiers, B., Mallants, D., Batelaan, O., Gedeon, M., Huysmans, M., Dassargues, A., 2012b. The usefulness of CPTs for deterministic spatially heterogeneous, large-scale aquitard parameterisation. In: Oswald, S.E., Kolditz, O., Attinger, S. (Eds.), Models - Repositories of Knowledge, IAHS Publ. 355: 41-47. Proceedings ModelCare 2011, Leipzig, Germany, 18-22 September 2011. - ISBN 978-1-907161-34-6.

Rogiers, B., Beerten, K., Smeekens, T., Mallants, D., Gedeon, M., Huysmans, M., Batelaan, O., Dassargues, A., 2013a. Derivation of flow and transport parameters from outcropping sediments of the Neogene aquifer, Belgium. *Geologica Belgica* **16**(3): 129-147.

Rogiers, B., Winters, P., Huysmans, M., Beerten, K., Mallants, D., Gedeon, M., Batelaan, O., Dassargues, A., 2013b. High resolution hydraulic conductivity logging of borehole cores using air permeability measurements. *Hydrogeology Journal*, submitted.

Rogiers, B., Vienken, T., Gedeon, M., Batelaan, O., Mallants, D., Huysmans, M., Dassargues, A., 2013c. Multi-scale aquifer characterization and groundwater flow model parameterization using direct push technologies. *Environmental Earth Sciences*, submitted.

Rubin, Y., Chen, X., Murakami, H., Hahn, M. 2010. A Bayesian approach for inverse modeling, data assimilation, and conditional simulation of spatial random fields. *Water Resources Research* **46**(10): 1–23. doi:10.1029/2009WR008799

Schulze-Makuch, D., 2005. Longitudinal dispersivity data and implications for scaling behavior. *Ground Water* **43**(3) 443–456.

Šimůnek, J., Sejna, M., van Genuchten, M.T., 2005. HYDRUS-1D, version 4.14, code for simulating the one-dimensional movement of water, heat, and multiple solutes in variably saturated porous media, Tech. rep., University of California Riverside, 2005.

Stedinger, J. R., Vogel, R. M., Lee, S. U., Batchelder, R., 2008. Appraisal of the generalized likelihood uncertainty estimation (GLUE) method. *Water Resour. Res.* **44**(12), W00B06, doi, 101029/.

ter Braak, C. J. F., 2006. A Markov Chain Monte Carlo version of the genetic algorithm differential evolution: Easy Bayesian computing for real parameter space, *Stat. Comput.* **16**: 239–249, doi:10.1007/s11222-006-8769-1.

ter Braak, C. J. F., Vrugt, J. A., 2008. Differential evolution Markov chain with snooker updater and fewer chains, *Stat. Comput.* **18**: 435–446. doi:10.1007/s11222-008-9104-9.

Tonkin, M., Doherty, J., 2009. Calibration-constrained Monte Carlo analysis of highly parameterized models using subspace techniques. *Water Resources Research* **45**: 1–17. doi:10.1029/2007WR006678

Vrugt, J. A., ter Braak, C. J. F., Clark, M. P., Hyman, J. M., Robinson, B. A., 2008a. Treatment of input uncertainty in hydrologic modeling: Doing hydrology backward with Markov chain Monte Carlo simulation, *Water. Resour. Res.* **44**(12): W00B09, doi:10.1029/2007WR006720.

Vrugt, J. A., Braak, C. J. F., Gupta, H. V., Robinson, B. A., 2008b. Equifinality of formal (DREAM) and informal (GLUE) Bayesian approaches in hydrologic modeling? *Stochastic Environmental Research and Risk Assessment* **23**(7): 1011–1026. doi:10.1007/s00477-008-0274-y

Vrugt, J. A., ter Braak, C. J. F., Diks, C. G. H., Higdon, D., Robinson, B. A., Hyman, J. M., 2009. Accelerating Markov chain Monte Carlo simulation by differential evolution with self-adaptive randomized subspace sampling, *Int. J. Nonlin. Sci. Numer. Simul.* **10**(3): 273–290.

Winton, C., Pettway, J., Kelley, C. T., Howington, S., Eslinger, O. J., 2011. Application of Proper Orthogonal Decomposition (POD) to inverse problems in saturated groundwater flow. *Advances in Water Resources* **34**(12): 1519–1526. doi:10.1016/j.advwatres.2011.09.007

Zheng, C., Wang, P. P., 1999. MT3DMS: A Modular Three-Dimensional Multispecies transport Model for Simulation of Advection, Dispersion, and

Chemical Reactions of Contaminants in Groundwater Systems; Documentation and User's Guide. US Army Corps of Engineers, Contract report SERDP-99-1, December 1999

Zheng, C., Hill, M. C., Hsieh, P. A., 2001. MODFLOW-2000, The U.S. Geological Survey Modular Ground-Water Model—User Guide to the LMT6 Package, the Linkage with MT3DMS for Multi-Species Mass Transport Modeling, U.S. Geological Survey Open File Report 01-82, Denver, Colorado, 2001

Chapter 10

General conclusions and recommendations

10.1. Introduction

In this work, a methodology is developed for efficient multi-scale subsurface characterization and integration of the gathered data in a stochastic regional groundwater flow and solute transport modelling approach. Different kinds of aquifer characterization techniques were used to cover the centimetre- to the kilometre-scale. To make optimal use of common or easily gathered secondary data, a set of efficient tools for data calibration and interpretation was employed. The integration of the different outcrop analogue, borehole and direct-push datasets was achieved using different innovative methods and techniques.

Figure 10.1 shows a flow chart illustrating how the different types of data were analyzed and interpreted in the different chapters of this thesis. Chapter 2 and 3 document development and application of a methodology for the quantification of hydraulic conductivity (*K*), its heterogeneity, as well as small-scale anisotropy and dispersivity on outcrops, using a hand-held air permeameter and grain-size analyses. The results obtained are compared to borehole core *K* data representative for the subsurface in Chapter 4. This borehole core *K* data is integrated with borehole core air permeameter measurements in Chapter 5, and used to develop a probabilistic data-driven model to predict *K* in function of grain size data in Chapter 6. The results of Chapter 5 and 6 are integrated with the direct push datasets in Chapter 8, to parameterize the updated groundwater flow model. Chapter 7 investigates the use of model-based clustering for cone penetration test (CPT) data, and demonstrates that the top of the Kasterlee Clay aquitard, the only lithostratigraphical boundary used in the upper ~40 m of the updated groundwater flow model, can be mapped with an automated approach. The dispersivity results from Chapter 3 are finally integrated with the groundwater flow model in Chapter 8, to quantify the uncertainty on groundwater flow, and indirectly on solute transport in the Neogene aquifer at Mol/Dessel, as described in Chapter 9.

The main scientific outcomes of the different chapters are recapitulated in section 10.2, and our findings concerning the different objectives are provided as well. Recommendations and future challenges are discussed in section 10.3.

10.2. Main scientific outcomes

Below, the main conclusions of the different chapters are recapitulated:

Chapter 2

Quantifying outcrop hydraulic conductivity and its spatial variability

We demonstrated that the handheld air permeameter is an efficient and accurate tool to characterize *K* and its spatial patterns *in situ* on outcrop sediments. It allows quantification of spatial variability in *K* on scales that are not normally feasible by classical core-based techniques, and it requires very little time. Moreover, equivalent outcrop-scale *K* tensors can be derived, and proved to be very informative. Such information can be valuable as prior information for use in geostatistical models, quantification of relative differences between lithostratigraphic units as well as anisotropy for groundwater flow modelling.

Validation of the air permeability (k_a)-based *K* estimates was demonstrated by comparison with constant head lab tests on 100 cm³ cores. The results show that the K estimates are reasonably accurate in the 10^{-8} to 10^{-3} m/s range.

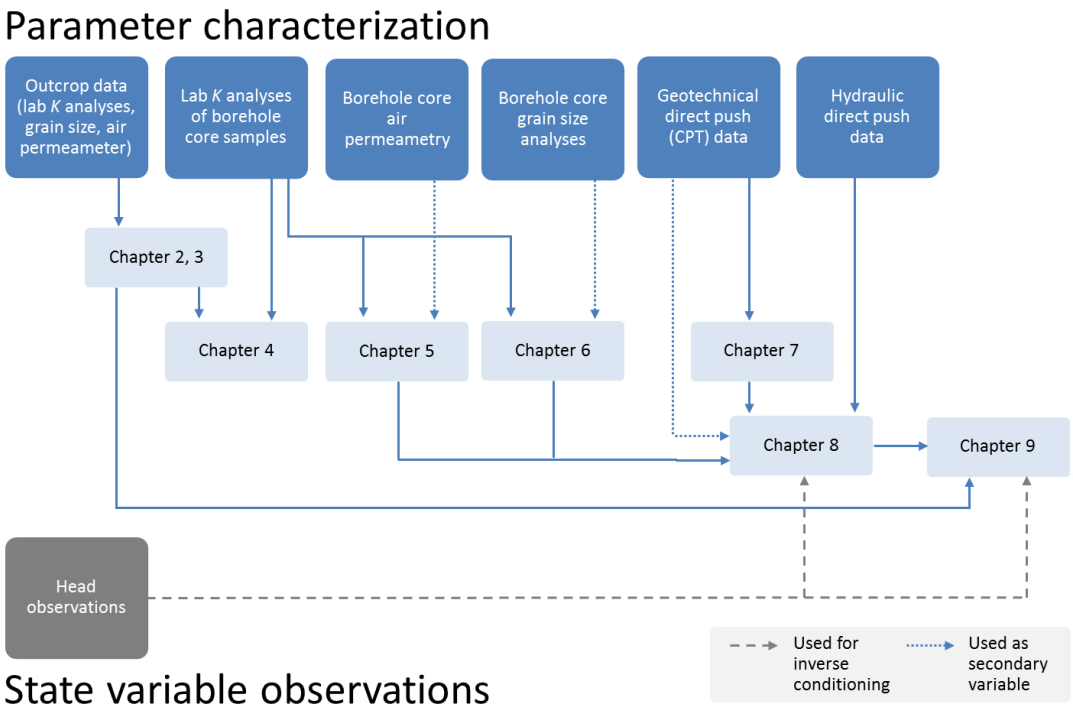

Figure 10.1: Flow chart of the different datasets described in this thesis, and the sequence of chapters in which they are used.

Chapter 3
Derivation of flow and transport parameters from outcropping sediments of the Neogene aquifer, Belgium
The methodology developed in Chapter 2 was applied to a larger set of outcropping sediments from the Neogene aquifer. The maximum range of *K* values obtained for one outcrop ranged from less than one order of magnitude to more than four orders of magnitude. This clearly indicates the presence of lithostratigraphical units with an important small-scale heterogeneity as well as very homogeneous units. Investigating spatial correlation for the different measurement grids showed that, at the considered outcrop scales, hardly any correlation exists for several formations, while others display a significant increase of the spatial variability with distance. The contrasts in heterogeneity between the different formations are very well illustrated by means of the experimental variography.
The methodology from Chapter 2 was extended by estimating the outcrop-scale dispersivity with particle tracking in combination with an equivalent random walk approach. Most of the calculated longitudinal dispersivities fit well within literature data, given the scale of the outcrops. A few outliers were obtained for outcrops with contrasting lithologies, where a dual domain approach might be more appropriate at the very small scale. The ratio between longitudinal and transverse dispersivities seems to be rather high, but falls in the range reported in literature. This builds confidence for the use of numerically calculated dispersivity values from outcrop sediments within larger-scale solute transport models.

Chapter 4
Testing outcrop hydrogeological parameters with independent borehole data
The comparison between outcrop and borehole core *K* values revealed significant differences between both datasets. Such differences are believed to be induced mainly by weathering, different palaeoenvironmental conditions (resulting in sorting or grain size differences) and differential compaction, and can be corrected for, as was demonstrated on the basis of a linear model. Hence, outcrop information can be used for building better stratigraphic models including determination of spatial structure by variogram fitting for further use in geostatistical simulations. Moreover, the relative variability in K values with similar coefficients

of variation for borehole and outcrop K, and the derived anisotropy values are very useful to get a more complete understanding of the heterogeneity within a heterogeneous aquifer.
In most cases the semivariograms for the outcrop and borehole data are however compatible. Spatial correlation (*i.e.* increasing semivariance with distance) is present in most cases, either in the outcrop or borehole data, or both, but pure nuggets are found as well. Next to the quantitative information tested in this paper, information about facies geometry, like alternating clay and sand layers, cannot be revealed easily using available in situ methods, and represents very important qualitative knowledge obtained from outcrops, that contributes considerably to the development of a conceptual hydrogeological model.

Chapter 5
High resolution hydraulic conductivity logging of borehole cores using air permeability measurements
We demonstrated that the hand-held air permeameter is also a very efficient cost-effective tool to obtain high-resolution information on *K* from borehole cores or borehole core slabs. The quantified measurement error proved to be small compared to the *K* variability. Calibration with independent lab-based *K* measurements improved the predictive capacity of existing equations in literature considerably. Based on a 5-cm measurement interval, the air permeability-based *K* values revealed considerable small-scale spatial variability on a set of cores of the Neogene aquifer, with an overall range in *K* between 10^{-10} and 10^{-3} m/s.
Spatial interpolation using the site-specific air permeability calibration with linear mixed effects models as secondary variable in an ordinary co-kriging approach proved to be reasonably accurate, based on a full leave-one-out cross-validation with an R^2 of 0.61. Especially the interpolated *K* estimates of the thin clay lenses improved drastically. Finally, a comparison of the interpolated high-resolution K_h and K_v profiles revealed that at the 100 cm^3 sample scale, anisotropy is obvious in certain sections, which was not evident from an analysis of the primary lab-based *K* dataset alone.

Chapter 6
Estimation of hydraulic conductivity and its uncertainty from grain-size data using GLUE and artificial neural networks
It was shown that using grain-size distributions together with linear and non-linear data-driven modelling gives strong improvement over traditional non-site-specific methods for predicting *K*. Multiple linear regression (MLR) combined with principal component analysis is found to be reasonably accurate for predicting hydraulic conductivity on the basis of grain-size data. Up to eight cumulative grain-size fractions were included in the analysis; all except of the two largest fractions contributed to explaining the variability in *K* data. This shows that nearly the entire particle-size distribution curve contains useful information to be included to build predictive *K* models.
The non-linear data-driven modelling was performed by combining artificial neural networks (ANN) and generalized likelihood uncertainty estimation (GLUE). This GLUE-ANN approach provided a greater accuracy in predicted *K*, and considerably smaller prediction intervals, with equal reliability. The GLUE-ANN model is thus preferred over the MLR model, especially in the framework of site specific stochastic groundwater flow modelling where small-scale variability in *K* is important. The predictions made using this model complete the other available *K* datasets for the Neogene aquifer at Mol/Dessel, so that a continuous vertical *K* profile is obtained.
To validate the developed models, a leave-one-out cross-validation was performed. The GLUE-ANN model proved to be more accurate than the MLR model, and the 95% prediction uncertainty range was reduced on average by half an order of magnitude.

Chapter 7
High-resolution hydrostratigraphic characterization of a heterogeneous sedimentary aquifer using cone penetration data.
In this Chapter, we have shown that model-based soil behaviour type (SBT) classification of CPT data can be useful for regional lithostratigraphical mapping. The obtained SBTs provide more detailed information than those obtained with frequently used deterministic unsupervised clustering algorithms like k- and x-means clustering. Moreover, the obtained classification honours more the intrinsic classes within the data, in contrast to the classical

literature SBT classification charts. These findings are obtained by looking at the multivariate sediment properties from cored boreholes, in function of the SBTs, or by studying the spatial distribution of the obtained classes. The use of stratigraphical depth for clustering proved to be useful for the presented case study, and is recommended for geologically layered sites with unconsolidated sediments.
We also proposed a methodology for automated lithostratigraphical mapping using site-specific SBTs, which was applied to map the top of the Kasterlee Clay aquitard within our study area. Comparison with the more traditional, time-consuming, manually interpreted results suggest that this methodology is highly efficient, and could at least be used to support manual interpretation based on literature SBT classifications that provide indications on lithology, but lack information on the true typology of the data.

Chapter 8
Multi-scale aquifer characterization and groundwater flow model parameterization using direct push technologies
We demonstrated that direct push technologies including CPTs are very much suited for aquifer characterization from the local to the regional scale. Moreover, we have illustrated that standard cone penetration test data can be used to estimate K if calibration data is available, allowing to map aquifer and aquitard properties in 3D. The use of hydraulic direct push tools is more time consuming than CPT testing, but a higher accuracy can be obtained with similar data resolution. Both types of direct push data were integrated together with available borehole data in a regional groundwater flow model. We considered all hydraulically obtained data as the primary variable, while the geotechnical data was considered as a secondary variable. A non-stationary geostatistical approach was used to condition the model parameter realizations. After local optimization, and the selection of the best out of 20 realizations, a considerable increase in model performance was obtained, compared to the reference case. Therefore, the use of effective direct push technologies is highly recommended for shallow aquifer heterogeneity characterization.

Chapter 9
Large-scale stochastic inversion using non-stationary variograms for transport uncertainty assessment of a near surface radioactive waste disposal
We combined Adaptive Metropolis and adaptive spatial resampling by block sampling to quantify the prediction uncertainty of the updated groundwater model introduced in Chapter 8. The corresponding groundwater flow fields were further used to estimate the uncertainty of solute transport. At first sight, the combination of AM and ASR by block sampling seems not to be very efficient for McMC sampling with the forward model (the groundwater flow model developed in Chapter 8). The results of using an ISR chain without adaptivity indicate that the adaptivity of the φ parameter (the fraction of constant nodes in the ISR/ASR method) might be too high, and additional tests should be performed to verify this. However, running the algorithm in optimization mode seems to be very efficient, and provides an alternate way for exploring the parameter space and approximating the posterior.
Despite the effect of numerical dispersion, the steady-state finite difference advection scheme solutions suggest that the approach with homogeneous layers in the reference model results in conservative estimates of the dilution rate, compared to the inclusion of subsurface heterogeneity. The most pessimistic parameter set that is considered part of the posterior should however be chosen to obtain a more robust probabilistic conservative estimate. The few transient simulations with the method of characteristics however show that inclusion of physical dispersion might lead to a decrease of the dilution rate. This should be investigated in more detail. The size of the solute plume, on the other hand, seems to be considerably underestimated by the use of homogeneous hydrogeological units, yet finer numerical discretization and the use of Eulerian-Lagrangian methods are recommended for obtaining more accurate estimates. The use of local-scale dispersion in the updated model that accounts for subsurface heterogeneity, and the use of macroscopic dispersion in the homogeneous reference model however lead to a similar result.

With respect to the different objectives, the following conclusions can be drawn:

Overall goal
Quantify the effect of multi-scale heterogeneity on groundwater flow and solute transport in the Neogene aquifer at the Mol/Dessel nuclear zone

Primary objective 1
Hydraulic conductivity heterogeneity: sources, characterization, effects

Secondary objective a
Quantify spatial variability
For quantification of the spatial variability within the Neogene aquifer, we made use of different datasets collected at different spatial scales. The similar small-scale measurement supports of most of these datasets facilitated their geostatistical integration. For the final geostatistical *K* realizations for the groundwater flow model, the borehole and direct push datasets were upscaled, and analyses of spatial variability was performed for these larger spatial supports. The spatial variability within the outcrops was used in a different way, and was inserted in the final modelling of solute transport through the estimates of physical dispersivity, both at the local and macroscopic scale.

Secondary objective b
Investigation of small-scale anisotropy and dispersivity
The existence of small-scale anisotropy was confirmed both by the derivation of equivalent *K* tensors in the outcrop investigations, as well as the separate treatment of K_h and K_v when integrating the borehole core air permeameter measurements with the laboratory analyses. By using the resulting K_h and K_v borehole datasets for calibration of the CPT data, this small-scale anisotropy is translated to the secondary data, and is accounted for in the upscaled datasets used for the groundwater flow model parameterization.
As mentioned above, dispersivity was estimated from the outcrop investigations, and used in the transport model in Chapter 9 for accounting for realistic small-scale *K* variability.

Secondary objective c
Study clay lens geometry and occurrence
The lack of knowledge on the thin clay lenses within the Kasterlee Clay aquitard was solved partly by studying the outcrop analogue at Heist-op-den-Berg, and calibration of the CPT data with the high-resolution borehole core *K* logs. The outcrop analogue suggests that the horizontal continuity of these lenses is at least several tens of meters. Moreover, the resemblance with the bottom of the Poederlee Formation at Kasterlee, where several clay lenses seemed to form a continuous connected network at least along 150 m of the tunnel excavation, provides confidence in the use of harmonic and arithmetic means for upscaling *K*.
The use of the boundary energy shape characteristic of the CPT friction ratio to enhance the detection of these thin lenses was only possible given the high-resolution borehole core *K* logs. This indicates that high quality calibration data is very much needed for estimation of *K* from CPT data when small-scale heterogeneity is present. The use of hydraulic direct push tools however provides an alternative for gathering such high-resolution data.

Primary objective 2
Condition flow and transport models on the available data

Secondary objective d
Data-driven identification of hydrogeological units
The idea to work with data-driven identification of hydrogeological units was mainly to avoid sharp boundaries within the simulated *K* fields of the groundwater flow model, that can be criticized with alternate lithostratigraphical interpretations, or the possible existence of more gradual transition zones between layers. The only lithostratigraphical boundary that was finally used within the upper 40 m of the Neogene aquifer within the study area is the top of the Kasterlee Clay aquitard, for which we showed that it can be mapped in an objective automated way by making use of site-specific SBT classification in Chapter 7. Omission of the other lithostratigraphical boundaries was achieved by using the non-stationary geostatistical approach, which can be seen as a continuously changing data-based lithostratigraphical column.
For the lower part of the Neogene aquifer, such an approach was however not possible, and the lithostratigraphical boundaries were taken as implemented in the reference groundwater flow model. Nevertheless, the parameterization of

this domain was done continuously, based on 1D data from the grain-size-based *K* predictions and the deep Dessel-5 borehole data.

Secondary objective e
Predict K from secondary data
Three types of secondary data were used in this work for the estimation of *K*: air permeability data, CPT data, and grain size distributions. For each of these data types, methods to estimate *K* exist in literature. However, we tested each time the site-specific calibration of the secondary data, repectively with a linear mixed effects model, the combination of linear models with a correction based on the friction ratio boundary and an upscaling step, and a non-linear data-driven approach combining GLUE and ANNs. Site-specific calibration each time improved the accuracy considerably, and for the GLUE-ANN approach also the prediction uncertainty.

Secondary objective f
Condition flow and transport models
Conditioning of the groundwater flow model was performed making use of a non-stationary multivariate geostatistical framework, in which all borehole and hydraulic direct push data was used as a primary variable, and the geotechnical direct push *K* estimates (CPTs) as secondary variable. This approach was applied for the upper ~40 m of the Neogene aquifer. As less data is available for the deeper part of the aquifer, a 1D data-based parameterization was used for that part of the model.
The transport model is indirectly conditioned on all this data as well, by making use of the groundwater flow model flow fields. Moreover, realistic estimates of dispersivity based on the outcrop investigations were used as well.

Primary objective 3
Quantify uncertainty and assess the effect of dispersion

Secondary objective g
Quantify prediction uncertainty
The uncertainty on groundwater flow and the corresponding solute transport is quantified by using a Markov chain Monte Carlo approach, combining adaptive Metropolis (AM) and adaptive spatial resampling (ASR) by block sampling. The tested McMC sampling with the updated groundwater flow model seems not to be very efficient, due to the systematically decreasing likelihood in the chain, based on a very limited number of forward model runs. This is probably due to too large adaptivity of the number of resampled nodes in the ASR algorithm. Further testing should be performed to clarify this. Additionaly, we used the developed algorithm in optimization mode, and use different chains to perform inverse conditioning, and keep some variability at the same time. As convergence of the chains could not be demonstrated, the selected samples of the flow solutions only provide an approximation of the posterior.

Secondary objective h
Introduce dispersivity coefficients in transport model
Dispersion is introduced in the transport simulations making use of the method of characteristics (MOC). This was done to provide more accurate analyses of the predicted solute plume, as the finite difference-based solutions used for the model ensembles are subject to considerable numerical dispersion. Dispersivity was estimated based during the outcrop investigations, and projected to the required spatial scale using the dispersivity-scale relationship derived from literature data. For the updated model accounting for subsurface heterogeneity, a local-scale dispersion was used; for the homogeneous reference model, macroscopic dispersion was derived as well. The resulting deterministic spatially distributed dispersivities seem much more realistic than the use of a single estimated value, as it accounts for the varying degrees of heterogeneity encountered within the Neogene aquifer.

Secondary objective i
Provide feedback for the cAt program
The stochastic flow model ensemble results combined with the steady-state finite difference advection solution suggests that the dilution rate obtained from the reference model with homogeneous *K* fields are conservative. The effect of dispersion on the dilution rate should however be investigated in more detail. On the other hand, the plume volume obtained with this homogeneous model is underestimated by a factor 2, according to the finite difference solution. For obtaining accurate probabilistic results, simulations with finer numerical discretization are however recommended, even with the MOC advection solution. The dilution rate however is much less affected by dispersion (both numerical and physical), as it considers

the average concentration within the upper aquifer, close to the solute source.
The obtained flow solution ensembles are however at least as important as the transport investigations, as they can learn us a lot on the hydrogeological system, and further processing of these results is recommended.

In general, we conclude that multi-scale aquifer characterization by combining outcrop analogue, borehole and direct push investigations is a very effective way to inform groundwater flow and solute transport models on a regional scale, with an emphasis on aquifer heterogeneity and small-scale variability. The consistent use of small-scale measurement supports facilitates the integration of different datasets, and the availability of different direct push technologies allows to gather such data at a high resolution in an effective way. The more traditional use of cored boreholes of course delivers the highest data quality, and using a hand-held air permeameter to characterize the cores again leads to high resolution data in a very effective way.

10.3. Recommendations and future challenges

10.3.1. The cAt programme

Given the influence of numerical dispersion in the solute transport simulations in Chapter 9, and the limited testing with an Eulerian-Lagrangian advection solution, it is recommended to perform further calculations using the latter method, and if possible to use it for the flow model ensembles. Moreover, other solution schemes should be tested as well. Refinement of the numerical grid for the transport calculations is however recommended in any case. If the grid can be refined considerably, the finite difference steady-state solution might provide more accurate results, still with the benefit of lower CPU-time, but this should be tested. Another approach would be to make use of a pure Lagrangian random walk method for simulation of the solute transport, but such a code that is compatible with MODFLOW is currently not available yet, and some code development would be required.
The presence and geometry of the high K coarse Upper Mol sands that are characterized with the different hydraulic direct push tests should be further explored. This lithostratigraphical unit should be clearly mapped separately, and at least considered to be included in the reference model with homogeneous layers, as it certainly influences the groundwater flow beneath the eastern tumulus (solute source) and the flux at the source location is determining for the dilution rate.
As pumping test data is available at several locations within the study area, it would be interesting to use this data to expand the inverse conditioning of the groundwater flow model. This could easily be achieved by extracting part of the geostatistical realization around the pumping test sites, and running a numerical model for reproducing the pump test. The match with the observed data can then be integrated in the likelihood function, and will exert a larger-scale control on the geostatistical realizations.
The parameter characterization versus state variable observation is somehow out of balance (Figure 10.1), and the integration of different state variables in the modelling approach is recommended to further reduce the uncertainties. This may include the use of groundwater age tracing, in order to find the correct balance between recharge and hydraulic conductivity, groundwater flux measurements, the use of temperature as a groundwater tracer, *etc.* Moreover, to validate the dispersivity estimates, the use of a single, or even multiple tracer tests is recommended.

10.3.2. The international research context

Given the many different subjects treated in this work, a number of recommendations for future research can be made:
The outcrop studies were only performed on a limited set of outcrops, with mostly one outcrop per lithostratigraphical unit. Of course, the representativity of a single outcrop is questionable. Moreover, the comparison between outcrop and subsurface data highlighted the existence of systematic bias between the datasets, although the relative differences between the outcrops proved to be useful. Therefore, future research should aim at the systematic characterization of different outcrops of a single lithostratigraphical unit, to obtain more representative results. To tackle the discrepancy with subsurface data, a more systematic study of K with depth, and the relationship with porosity, density, differential compaction and weathering is recommended. A heterogeneous aquifer with dipping layers is

however not the place to solve such a problem, as the heterogeneity overprints the targeted effects and more homogeneous sediments should be targeted.

For the characterization of borehole cores with a hand-held air permeameter, we recommend to further investigate the combined analysis of K_h and K_v datasets, but the issue of non-colocated K_h and K_v core samples should be tackled. The use of pseudo-variography might solve this problem, and make the approach more generic, and applicable for K_h and K_v data that is not gathered systematically at small distances from each other. For future applications of the methodology, it is also recommended to perform the air permeameter measurements as quickly as possible after retrieving the cores, and removal and drying of the core slabs. This would possibly increase the accuracy of the air permeability results, and the calibration effort might be a lot smaller.

The GLUE-ANN methodology that was developed for estimating K based on grain size distributions provides a generic non-linear data-driven modelling tool. The approach can easily be extended with additional parameters like organic matter and carbonate content, or soil porosity and density. A hierarchical set of models could also be derived from a given incomplete dataset, to provide consistent predictions in case of different data availability. The number of target variables can also be increased, and the statistical dependence between the target variables from the training data would be honoured in the predictions as well. Extrapolation should however be avoided, as the ANN approach is very flexible, and might provide erroneous estimates for data that is too different from the data the model was trained with.

The automated lithostratigraphical mapping based on model-based clustering of CPT data was only tested on a single horizon within the lithostratigraphical column. Further research should focus on the joint mapping of different horizons and layers, possibly in a sequential approach. Moreover, to make the identification of the horizons more robust, constraints could be placed on the differences between identified depths at nearby CPTs, including prior knowledge on the layer geometries.

The estimation of K based on CPT data was performed using available cored borehole data. Future research should focus on omitting the need for such data, and replacing it with high-resolution hydraulic direct push data. A few well-chosen hydraulic direct push test sites could then complement a regional CPT campaign, and allow for high-resolution, high-detail multi-scale shallow aquifer characterization without the need for expensive drilling.

Appendix A

R script with examples of GLUE-ANN analyses

```
##############################################################################
### GLUE-ANN parameters + examples (Rogiers et al. 2012)                   ###
##############################################################################
# A dataframe has to be provided with the independent variable(s), and a vector
# with the dependent variable. A list of parameters has to be set to define the
# a priori distributions of the stochastic parameters, and fixed values for the
# ones considered to be constant.
# Be sure to load the additional GLUE-ANN functions into R before trying
# the examples provided below. The AMORE package should be installed and loaded:
install.packages('AMORE')
library('AMORE')

### Parameter list ###########################################################
# A list of all required parameters with their default values.

  ### Pre-processing parameters ##############################################
  # Orthogonalise input data; options are: 'Cov', 'Cor', and 'No'
    orthogonal <- 'Cor'
  # Number of input variables, provide vector for stochastic treatment
    variableNumber <- c(1:ncol(inputData))
  # T: Select always most variance explaining PCs; F: Choose PCs randomly
    dimRed <- T
  # Rescale output to 0-1 range
    rescale.output <- T

  ### ANN parameters ###########################################################
  # Enable early stopping
    ES <- T
  # Part of the data used for early stopping
    randomESPart <- 0.5
  # Early stopping samples should have a similar mean as the training samples
    nonEqualMeans <- T
  # Maximum difference between early stopping and training observation means
    maxMeanDiff <- mean(outputData)*0.2
  # maximum number of training cycles, provide a vector for stochastic treatment
    nCycles <- 100
  # Amount of hidden layers, provide vector for stochastic treatment
    HLrange <- c(1:1)
  # Amount of hidden nodes, provide vector for stochastic treatment
    HNrange <- c(1:10)
  # Maximum amount of hidden nodes per input variable
    maxHNPerVar <- Inf
  # Hidden layer activation function; options are 'tansig', 'sigmoid', 'purelin'
    hidden.layer <- 'tansig'
  # Output layer activation function; options are 'tansig', 'sigmoid', 'purelin'
    output.layer <- 'sigmoid'

  ### GLUE parameters ##########################################################
  # Number of models in the ensemble
    nSets <- 100
  # Number of random variates for weights based on Stedinger et al (2008)
    nStedinger <- 50
  # GLUE weighting procedure; options are 'MSE', 'MSEes' (for weights based on
  # the early stopping subset performance), 'NSEff' and 'Stedinger'
    weighting <- 'Stedinger'

  ### Cross-validation parameters ##############################################
  # Use leave-one-out cross-validation
    cv <- T

### Examples #################################################################

  ### Example of training and prediction for fictional data ####################
  # Create fictional data for x range [1,100]
    inputData <- data.frame(x=seq(1,100,0.5))
  # Dependent variable y = x^2
    outputData <- inputData$x ^ 2
```

```
  # add random noise to y, representing different sources of error
    outputData <- jitter(outputData, amount=mean(outputData)*0.2)
  # Run GLUE-ANN functions
    glue.ann.ensemble <- train.glue.ann(inputData, outputData, nCycles=500)
    training.results <- predict.glue.ann(inputData, glue.ann.ensemble)
  # Visualise results for training data
    plot(inputData$x, training.results$eMean, type='l', xlab='x', ylab=
    expression(y = x^2 + epsilon), ylim=c(-2000,13000))
    lines(inputData$x, glue.ann.ensemble$outputData, lty=2, col='red')
    lines(inputData$x, training.results$eQ025, lty=3, col='blue')
    lines(inputData$x, training.results$eQ975, lty=3, col='blue')
    legend('topleft', c('GLUE-ANN ensemble mean prediction','Observations',
    '95% prediction uncertainty'), lty=c(1,2,3),col=c('black','red','blue'),
    bty='n')
  # Predict new data
    newData <- data.frame(x=c(-100:200)+0.5)
    newData.predictions <- predict.glue.ann(newData, glue.ann.ensemble)
  # Visualise results of prediction
    plot(newData$x, newData.predictions$eMean, type='l', xlab='x', ylab=
    expression(y = x^2 + epsilon), ylim=c(-2000,13000))
    lines(newData$x, jitter(newData$x^2, amount=mean(outputData)*0.2) , lty=2,
    col='red')
    lines(newData$x, newData.predictions$eQ025, lty=3, col='blue')
    lines(newData$x, newData.predictions$eQ975, lty=3, col='blue')
    abline(v=c(0,100))
    text(50,13000,'Training data range')
    legend('bottomright', c('GLUE-ANN ensemble mean prediction','Observations',
    '95% prediction uncertainty'), lty=c(1,2,3),col=c('black','red','blue'),
    bty='n')

  ### Example of cross-validation with an R dataset ##########################
  # Read input and output data (ozone concentration in function of wind,
  # temperature and solar radiation)
    outputData <- na.omit(airquality)[,1]
    inputData <- na.omit(airquality)[,2:4]
  # Perform cross-validation
    cross_validation <- train.glue.ann(inputData, outputData, cv=T, nCycles=1000
    , nSets=10)
  # Visualisation
    plot(cross_validation$target, ylim=c(-100,200), xlab='Sample number', ylab=
    'Ozone')
    points(cross_validation$eMean, pch=2, col='red')
    lines(cross_validation$eQ025, lty=3, col='blue')
    lines(cross_validation$eQ975, lty=3, col='blue')
    legend('bottomright', c('Observations','GLUE-ANN ensemble mean prediction',
    '95% prediction uncertainty'), pch=c(1,2,NA), lty=c(NA,NA,3),col=c('black',
    'red','blue'),bty='n')
```

R functions for GLUE-ANN analysis with the AMORE package (Castejón Limas et al 2010)

```
##############################################################################
### GLUE-ANN - functions (Rogiers et al. 2011)                             ###
##############################################################################

### ANN training function ##################################################
nncalc <- function(Ptrain, Pval, targetTrain, targetVal, architecture, n, Stao=1
, es=T, hidden.layer="sigmoid", output.layer="purelin")
{
  neuralnetwork <- newff(n.neurons=architecture, learning.rate.global=1e-2,
  momentum.global=0.5, error.criterium = 'LMS', Stao = Stao, hidden.layer=
  hidden.layer, output.layer=output.layer, method="ADAPTgdwm")
  if(!es) { Pval <- NULL; targetVal <- NULL }
  capture.output(results <- train(neuralnetwork, Ptrain, Stao=Stao,  Pval=Pval,
  Tval=targetVal,targetTrain, error.criterium = 'LMS', report=T, show.step=1,
  n.shows=n ))
  return(results)
}

### Sampling parameter vectors function ####################################
resample <- function(x, n, ...) x[sample.int(length(x), size=n, ...)]

### Model ensemble training function #######################################
train.glue.ann <- function(inputData,outputData,orthogonal='Cor',variableNumber=
c(1:ncol(inputData)),dimRed=T,rescale.output=T,ES=T,randomESPart=0.5,
nonEqualMeans=T,maxMeanDiff=mean(outputData)*0.2,nCycles=100,HLrange=c(1:1),
HNrange=c(1:10),maxHNPerVar=Inf,hidden.layer='tansig',output.layer='sigmoid',
nSets=100,nStedinger=50,weighting='Stedinger',cv=F)
{
  ### Initialisation #######################################################
  parameters <- NULL
  parameters$orthogonal <- orthogonal
  parameters$variableNumber <- variableNumber
```

```
parameters$dimRed <- dimRed
parameters$rescale.output <- rescale.output
parameters$ES <- ES
parameters$randomESPart <- randomESPart
parameters$nonEqualMeans <- nonEqualMeans
parameters$maxMeanDiff <- maxMeanDiff
parameters$nCycles <- nCycles
parameters$HLrange <- HLrange
parameters$HNrange <- HNrange
parameters$maxHNPerVar <- maxHNPerVar
parameters$hidden.layer <- hidden.layer
parameters$output.layer <- output.layer
parameters$nSets <- nSets
parameters$nStedinger <- nStedinger
parameters$weighting <- weighting
parameters$cv <- cv
glue.ann <- NULL
cvPredictions <- NULL
glue.ann$parameters <- parameters
glue.ann$esNumber <- matrix(nrow=nrow(inputData), ncol=nSets)

### Training loop ############################################################
for(crossValidationSampleNumber in 1:ifelse(cv,nrow(inputData),1))
{
  if(cv)
  {
      inputDataModellingCV <- as.data.frame(
      inputData[crossValidationSampleNumber,])
      inputDataModelling <- as.data.frame(
      inputData[-crossValidationSampleNumber,])
      outputDataModellingCV <- outputData[crossValidationSampleNumber]
      outputDataModelling <- outputData[-crossValidationSampleNumber]
  } else {
      inputDataModelling <- inputData
      outputDataModelling <- outputData
  }
  names(inputDataModelling) <- c(1:ncol(inputDataModelling))
  if(cv) names(inputDataModellingCV) <- c(1:ncol(inputDataModellingCV))
  glue.ann$inputData <- inputDataModelling
  glue.ann$outputData <- outputDataModelling
  P <- cbind(inputDataModelling)
  if(orthogonal != 'No'){Ppca <- princomp(P, cor=ifelse(orthogonal=='Cor', T,
  F));P <- Ppca$scores}
  trans <- P
  glue.ann$P <- P
  for(i in 1:ncol(inputData)) trans[,i] <- (trans[,i]-mean(P[,i]))/sd(P[,i])
  if(rescale.output) outputDataModelling <- (outputDataModelling - (
  min(glue.ann$outputData)))/(((max(glue.ann$outputData))-(
  min(glue.ann$outputData))))
  if(rescale.output & cv) outputDataModellingCV <- (outputDataModellingCV -(
  min(glue.ann$outputData)))/(((max(glue.ann$outputData))-(
  min(glue.ann$outputData))))
  variableslist <- list();networks <- list(); esNumber <- NULL
  earlyStoppingMSE <- NULL; earlyStoppingCycle <- NULL; hn <- list()
  totalVar <- NULL; trainingMSE <- NULL
  for(j in 1:nSets)
  {
    if(cv) cat(paste('# Cross-validating sample',crossValidationSampleNumber,
    ' '))
    esDataNumbers <- sample(1:nrow(inputDataModelling),
    round(nrow(inputDataModelling)*randomESPart,0))
    while(nonEqualMeans)
    {
      esDataNumbers <- sample(1:nrow(inputDataModelling),
      round(nrow(inputDataModelling)*randomESPart,0))
      esMean <- mean(outputDataModelling[esDataNumbers])
      trainMean <- mean(outputDataModelling[-esDataNumbers])
      if(abs(esMean-trainMean) < maxMeanDiff) nonEqualMeans <- F
    }
    trainingDataNumbers <- c(1:nrow(inputDataModelling))[-esDataNumbers]
    if(!ES) {trainingDataNumbers <- c(1:nrow(inputDataModelling))
    esDataNumbers <- trainingDataNumbers}
    nvars <- resample(variableNumber,1)
    variables <- resample(c(1:ncol(inputDataModelling)),nvars)
    if(dimRed) variables <- c(1:length(variables))
    variableslist[[j]] <- variables
    hiddenLayers <- resample(HLrange,1, replace=T)
    hiddenNodes <- resample(HNrange,hiddenLayers, replace=T)
    for(i in 1:hiddenLayers) if(maxHNPerVar < hiddenNodes[i]/nvars) {
    hiddenNodes[i] <- resample((min(HNrange):maxHNPerVar),1, replace=T)}
    ncyc <- resample(nCycles,1)
    cat('# Training ANN nr. ',j,', with architecture: ',sep='')
    cat(c(nvars,hiddenNodes[1:hiddenLayers],1), sep='-')
    Ptrain <- as.data.frame(trans[trainingDataNumbers,variables]   )
    Pval <- as.data.frame(trans[esDataNumbers,variables])
    targetTrain <- outputDataModelling[trainingDataNumbers]
    targetVal <- outputDataModelling[esDataNumbers]
    networks[[j]] <- nncalc(Ptrain, Pval, targetTrain, targetVal,
```

```
      architecture=c(nvars,hiddenNodes,1), ncyc, Stao=10, es=ES, hidden.layer=
      hidden.layer, output.layer=output.layer)
      if(ES) esNumber[j] <- which.min(networks[[j]]$Merror[,2])
      predTrain <- sim(networks[[j]]$net, Ptrain)
      trainingMSE[j] <- mean((targetTrain-predTrain)^2)
      if(ES) predEarlyStopping <- sim(networks[[j]]$net, Pval)
      if(ES) earlyStoppingMSE[j] <- mean((targetVal-predEarlyStopping)^2)
      if(ES) totalVar[j] <- var(c(predTrain-targetTrain, predEarlyStopping-
      targetVal))
      if(!ES) totalVar[j] <- var(c(predTrain-targetTrain))
      if(ES) earlyStoppingCycle[j] <- which.min(networks[[j]]$Merror[,2])
      hn[[j]] <- hiddenNodes
      if(ES) cat(paste(' # Training MSE: ',round(trainingMSE[j],2),
      ', early stopping MSE: ',round(earlyStoppingMSE[j],2),', cycles: ',
      earlyStoppingCycle[j],'\n', sep=''))
      if(!ES) cat(paste(' # Training MSE: ',round(trainingMSE[j],2),
      ', cycles: ',earlyStoppingCycle[j],'\n', sep=''))
    }
    glue.ann$networks <- networks
    glue.ann$variables <- variableslist
    glue.ann$trainMSE <- trainingMSE
    if(ES) glue.ann$esMSE <- earlyStoppingMSE
    if(ES) glue.ann$totalMSE <- (1-randomESPart) * glue.ann$trainMSE +
    randomESPart * glue.ann$esMSE
    if(!ES) glue.ann$totalMSE <- glue.ann$trainMSE
    glue.ann$totalVar <- totalVar
    glue.ann$hn <- hn
    if(orthogonal != 'No') glue.ann$pca <- Ppca

    ### Predict cross-validated sample ########################################
    if(cv)
    {
      cvPrediction <- predict.glue.ann(inputDataModellingCV, glue.ann)
      cvPredictions$MSE[crossValidationSampleNumber] <- mean(
      (outputData[crossValidationSampleNumber] - cvPrediction$eMean)^2)
      cvPredictions$eMean[crossValidationSampleNumber] <- cvPrediction$eMean
      cvPredictions$eMedian[crossValidationSampleNumber] <- cvPrediction$eMedian
      cvPredictions$eMin[crossValidationSampleNumber] <- cvPrediction$eMin
      cvPredictions$eMax[crossValidationSampleNumber] <- cvPrediction$eMax
      cvPredictions$eQ025[crossValidationSampleNumber] <- cvPrediction$eQ025
      cvPredictions$eQ975[crossValidationSampleNumber] <- cvPrediction$eQ975
      cvPredictions$MLE[crossValidationSampleNumber] <- cvPrediction$MLE
      cvPredictions$target[crossValidationSampleNumber] <-
      outputData[crossValidationSampleNumber]
    }
  }
  if(cv){return(cvPredictions)} else {return(glue.ann)}
}

### Model ensemble prediction function ########################################
predict.glue.ann <- function(inputData, glue.ann,
weighting=glue.ann$parameters$weighting)
{
  names(inputData) <- c(1:ncol(inputData))
  if(glue.ann$parameters$orthogonal != 'No') inputData <-
  predict(glue.ann$pca, newdata=inputData) # orthogonalise
  for(i in 1:ncol(inputData)) inputData[,i] <-
  (inputData[,i]-mean(glue.ann$P[,i]))/sd(glue.ann$P[,i]) # standardise

  ### Predict model ensemble results for new data ##############################
  resultsMatrix <- matrix(nrow=nrow(inputData), ncol=glue.ann$parameters$nSets)
  if(nrow(inputData) > 1) {for(i in 1:glue.ann$parameters$nSets) {
  resultsMatrix[,i] <- sim(glue.ann$networks[[i]]$net,
  inputData[,glue.ann$variables[[i]]])}}
  if(nrow(inputData) == 1) {for(i in 1:glue.ann$parameters$nSets) {
  resultsMatrix[,i] <- sim(glue.ann$networks[[i]]$net, t(
  inputData[,glue.ann$variables[[i]]]))}}

  ### Construct GLUE weights ####################################################
  if(glue.ann$parameters$weighting=='MSEes'){weightsGLUE <- (max(glue.ann$esMSE)
  - glue.ann$esMSE)
  weightsGLUE[which(weightsGLUE < 0)] <- 0
  weightsGLUE <- weightsGLUE/sum(weightsGLUE)}
  if(glue.ann$parameters$weighting=='MSE'){weightsGLUE <- (
  max(glue.ann$totalMSE) - glue.ann$totalMSE)^2
  weightsGLUE[which(weightsGLUE < 0)] <- 0
  weightsGLUE <- weightsGLUE/sum(weightsGLUE)}
  if(glue.ann$parameters$weighting=='Stedinger')
  {
  se2 <- min(glue.ann$totalVar)
  n <- length(glue.ann$outputData)
    weightsGLUE <- exp((-n/2)* (glue.ann$totalMSE)/(se2) )
  }
  if(glue.ann$parameters$weighting=='NSEff')
  {
    weightsGLUE <- (1-(glue.ann$totalMSE/var(glue.ann$outputData)))
    weightsGLUE[which(weightsGLUE < 0)] <- 0
  }
```

```
  ### Weight model predictions ################################################
  predictions <- NULL
  predictions$MLE <- resultsMatrix[1,which.min(glue.ann$totalVar)]
  if(glue.ann$parameters$weighting != 'Stedinger')
  {
    predictions$eMean <- apply(resultsMatrix, 1, "wtd.mean", weights =
    weightsGLUE, normwt=T)
    predictions$eMedian <- apply(resultsMatrix, 1, "wtd.quantile", weights =
    weightsGLUE, probs = c(0.5), normwt=T)
    predictions$eQ025 <- apply(resultsMatrix, 1, "wtd.quantile", weights =
    weightsGLUE, probs = c(0.025), normwt=T)
    predictions$eQ975 <- apply(resultsMatrix, 1, "wtd.quantile", weights =
    weightsGLUE, probs = c(0.975), normwt=T)
    predictions$eMin <- apply(resultsMatrix, 1, "min")
    predictions$eMax <- apply(resultsMatrix, 1, "max")
    predictions$eResults <- resultsMatrix
  }
  if(glue.ann$parameters$weighting=='Stedinger')
  {
    resultsMatrix_sted <- matrix(nrow=nrow(inputData), ncol=
    glue.ann$parameters$nSets*glue.ann$parameters$nStedinger)
    for(i in 1:nrow(inputData))
    {
      for(k in 1:glue.ann$parameters$nSets)
      {
        errorVector <- rnorm(glue.ann$parameters$nStedinger,mean=0,sd=sqrt(se2))
        resultsMatrix_sted[i,(((k-1)*glue.ann$parameters$nStedinger)+c(
        1:glue.ann$parameters$nStedinger))] <- errorVector + resultsMatrix[i,k]
      }
    }
    weightsGLUE <- rep(weightsGLUE, each=glue.ann$parameters$nStedinger)
    predictions$eMean <- apply(resultsMatrix_sted, 1, "wtd.mean", weights =
    weightsGLUE, normwt=T)
    for(i in 1:nrow(inputData))
    {
      quantiles <- wtd.quantile(resultsMatrix_sted[i,], weights = weightsGLUE,
      probs = c(0.025, 0.5, 0.975), normwt=T)
      predictions$eMedian[i] <- quantiles[2]
      predictions$eQ025[i] <- quantiles[1]
      predictions$eQ975[i] <- quantiles[3]
    }
    predictions$eResults <- resultsMatrix_sted
    predictions$eMin <- apply(resultsMatrix_sted, 1, "min")
    predictions$eMax <- apply(resultsMatrix_sted, 1, "max")
  }

  ### Backtransform of predictions according to training data output ###########
  if(glue.ann$parameters$rescale.output)
  {
    outputRange <- max(glue.ann$outputData)-min(glue.ann$outputData)
    outputMin <- min(glue.ann$outputData)
    predictions$MLE <- predictions$MLE * outputRange + outputMin
    predictions$eMean <- predictions$eMean * outputRange + outputMin
    predictions$eMedian <- predictions$eMedian * outputRange + outputMin
    predictions$eQ025 <- predictions$eQ025 * outputRange + outputMin
    predictions$eQ975 <- predictions$eQ975 * outputRange + outputMin
    predictions$eMin <- predictions$eMin * outputRange + outputMin
    predictions$eMax <- predictions$eMax * outputRange + outputMin
    predictions$eResults <- predictions$eResults * outputRange + outputMin
  }
  return(predictions)
}
```

Appendix B

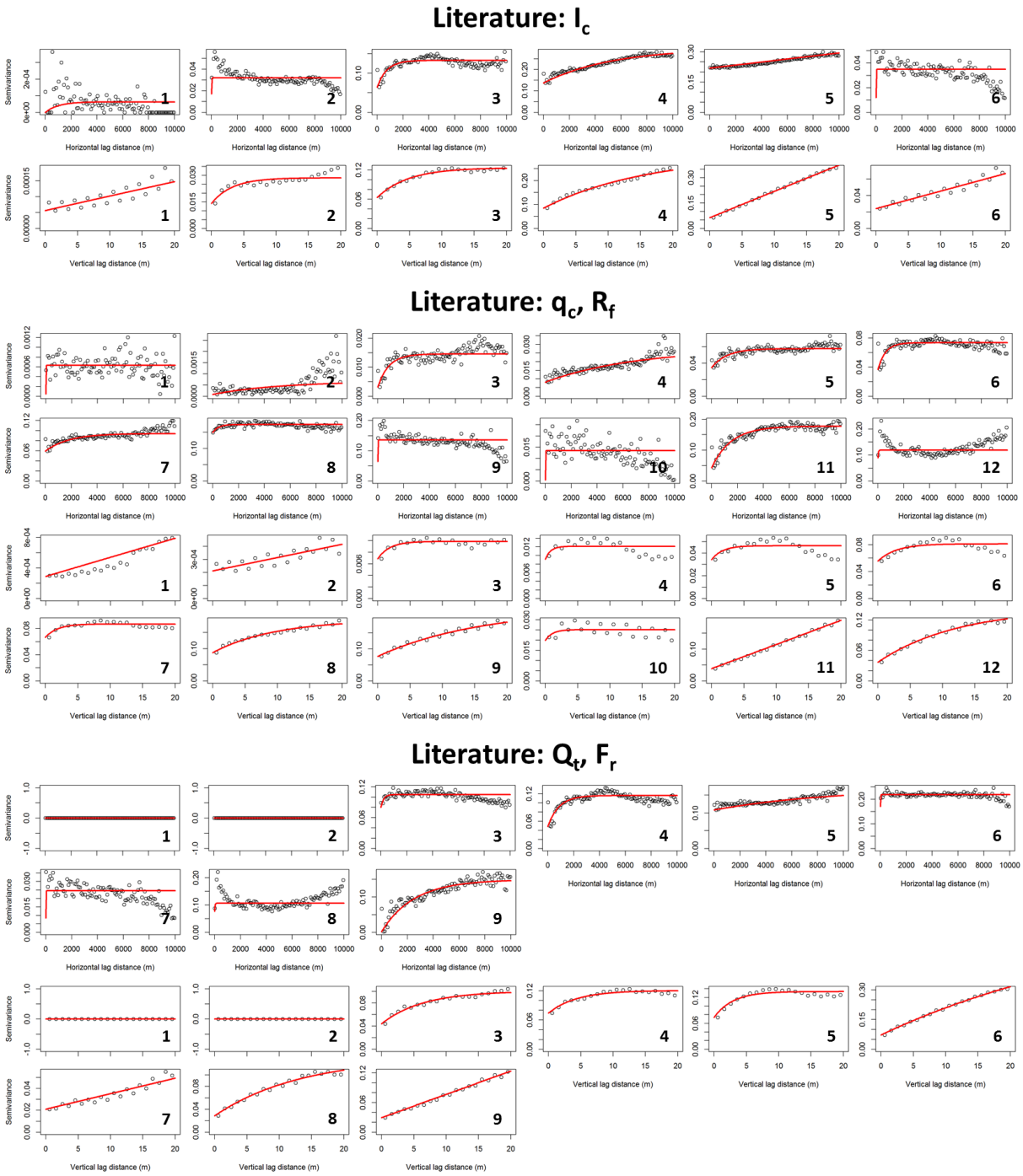

Figure A.1: Overview of the literature classification variograms.

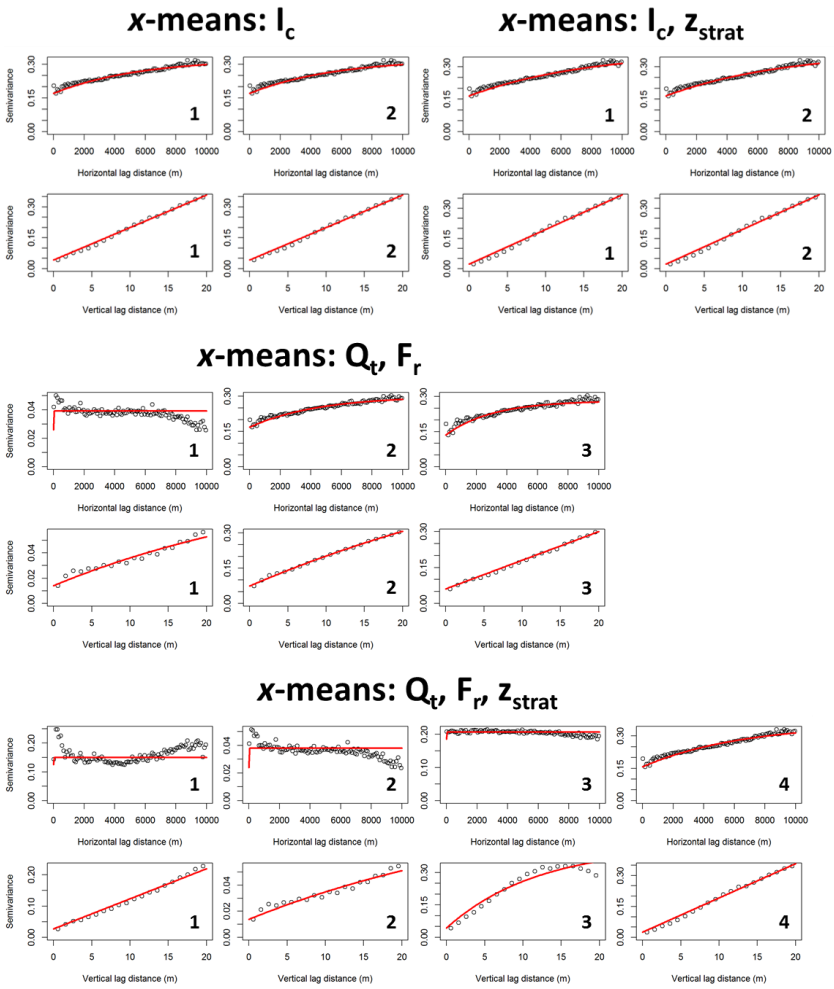

Figure A.2: Overview of the x-means classification variograms.

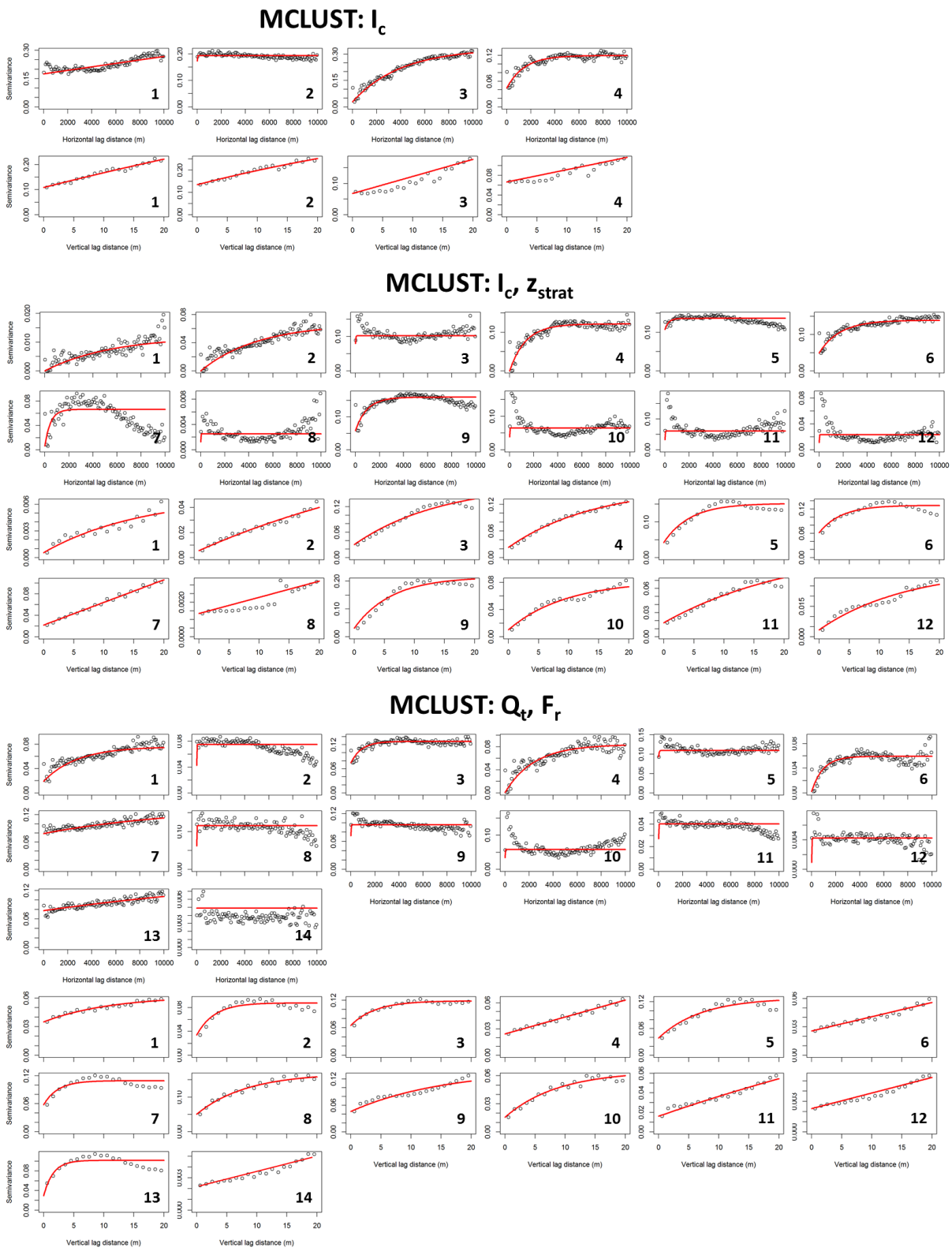

Figure A.3: Overview of the MCLUST classification variograms.

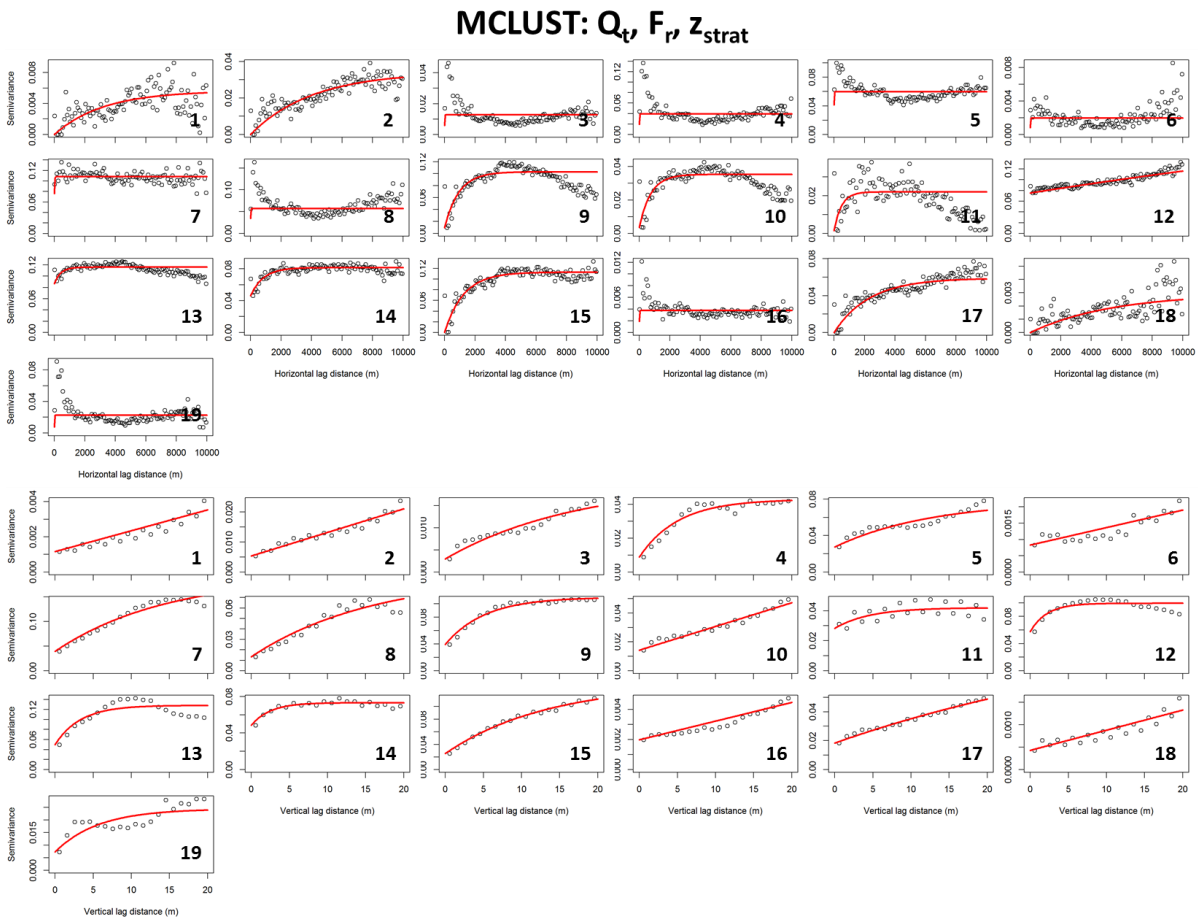

Figure A.3: (continued): Overview of the MCLUST classification variograms.

Curriculum vitae

Education

PhD

Nov 2009 - Present
Mol, Belgium
SCK•CEN & KU Leuven

Thesis
Multi-scale aquifer characterization: From outcrop analogue, direct-push and borehole investigations towards improved groundwater flow models.

Promoters
Prof. Dr. O. Batelaan; Dr. D. Mallants; Dr. M. Gedeon; Prof. Dr. M. Huysmans; Prof. Dr. Alain Dassargues

Followed courses
- Version control with subversion. Geert-Jan Bex (ICTS KU Leuven), February 26, 2013, Leuven.
- HPC introduction. - Martijn Oldenhof (ICTS KU Leuven), February 19, 2013, Leuven.
- Advanced R programming topics. - Jan Wijffels, BNOSAC, October 18-19, 2012, Leuven.
- Upgrade your written English. - Antoine Vermeire, April 19, May 3 & 4, 2012, Mol.
- Latex - introduction. - ICTS KU Leuven, March 29 & 30, 2012, Leuven.
- Data Mining in Practice. - Computer Sciences Department KU Leuven, February 6-7 , 2012, Leuven.
- Speedreading. - Bernard Lernout, October 11, 2011, Mol.
- Scientific Writing & Speaking. - Antoine Vermeire, March 2, 9, 21 & 28, 2011, Mol.
- Global Sensitivity Analysis Techniques for Probabilisitc Groundwater Modeling. - Srikanta Mishra, September 21, 2010, Valencia.
- 2nd Summer School on Flow and Transport in Porous and Fractured Media. - CNRS, August 16-28, 2010, Cargèse.
- Inverse Modelling in Earth and Environmental Sciences. - Jasper Vrugt & Sander Huisman, July 26-29, 2010, Leuven.
- Meet the expert in hydrology - Round tables among young and established scientists. - Jezus Carrera, May 5, 2010, Vienna.
- Short Course on Geostatistical Analysis of Environmental Data. - Pierre Goovaerts, March 8-12, 2010, Gainesville.
- Introduction to R. - ICTS KU Leuven & LStat, February 22-23, 2010, Leuven.
- Multivariate data analysis, with applications. - ICTS KU Leuven & LStat, February 8-10, 2010, Leuven.
- Hydrological modelling and uncertainty analysis. - Wouter Buytaert, January 18, 2010, Louvain-la-Neuve.

Internship

Sep 2008 - Mar 2009
Flemish Environment Agency,
Operational Water Management,
Groundwater Management
Brussels, Belgium

Subject

River restoration of the Zwarte Beek between Beringen and Diest: evaluation of the impact on the groundwater system.

Guidance
J. Lermytte

Master of Science in Geology

Sep 2007 - Jun 2009; magna cum laude
KU Leuven
Leuven, Belgium

Thesis
Interpretation of the heat-flow density in the deep wells Soumagne, Grand-Halleux and Havelange in Belgium, by means of a numerical coupled heat transport and groundwater flow model.

Promoters
Prof. Dr. N. Vandenberghe; Dr. M. Huysmans

Bachelor of Science in Geology, Major Subject: Option Physics

Sep 2004 - Jun 2007; cum laude
KU Leuven
Leuven, Belgium

Thesis
The meaning of heat-flow density measurements in the wells Grand-Halleux, Soumagne, Meer and Havelange.

Promoters
Prof. Dr. N. Vandenberghe; Dr. M. Huysmans

Peer reviewing

Journal of Hydroinformatics, 2011
Journal of Hydrology and Hydromechanics, 2013

Teaching experience

Hydrogeology teaching assistant, KU Leuven (2010 - 2013)

Master thesis jury member

Nafyad Serre Kawo, *Model Based Evaluation of the Protection Zones around the WaterGroep Public Drinking Water Well Field in Huiskens (Korbeek-Lo), Belgium.* Interuniversity programme: Master of Science in Physical Land Resources, Universiteit Gent - Vrije Universiteit Brussel. September 2013.

Grants and Awards

SCK•CEN PhD scholarship 2009-2013

List of publications

Papers

submitted

Rogiers B, Mallants D, Batelaan O, Gedeon M, Huysmans M, Dassargues A. High-resolution hydrostratigraphic characterization of a heterogeneous sedimentary aquifer using cone penetration data. *Water Resources Research*.

Rogiers B, Vienken T, Gedeon M, Batelaan O, Mallants D, Huysmans M, Dassargues A. Multi-scale aquifer characterization and groundwater flow model parameterization using direct push technologies. *Environmental Earth Sciences*.

Rogiers B, Winters P, Huysmans M, Beerten K, Mallants D, Gedeon M, Batelaan O, Dassargues A. High resolution hydraulic conductivity logging of borehole cores using air permeability measurements. *Hydrogeology Journal.*

in internationally reviewed scientific journals

Rogiers B, Beerten K, Smeekens T, Mallants D, Gedeon M, Huysmans M, Batelaan O, Dassargues A. 2013. The usefulness of outcrop analogue air permeameter measurements for analysing aquifer heterogeneity: testing outcrop hydrogeological parameters with independent borehole data. *Hydrol. Earth Syst. Sci.*, accepted. (*Hydrol. Earth Syst. Sci. Discuss.* **10**: 9689-9720).

Rogiers B, Huysmans M, Vandenberghe N, Verkeyn M. Demonstrating large-scale cooling in a Variscan terrane by coupled groundwater and heat flow modelling. *Geothermics*, accepted.

Rogiers B, Beerten K, Smeekens T, Mallants D, Gedeon M, Huysmans M, Batelaan O, Dassargues A. 2013. Derivation of flow and transport parameters from outcropping sediments of the Neogene aquifer, Belgium.*Geologica Belgica* **16**(3): 129-147.

Rogiers B, Beerten K, Smeekens T, Mallants D, Gedeon M, Huysmans M, Batelaan O, Dassargues A. 2013. The usefulness of outcrop analogue air permeameter measurements for analysing aquifer heterogeneity: Quantifying outcrop hydraulic conductivity and its spatial variability. *Hydrological processes*.

Boden S, Rogiers B, Jacques D. 2013. Determination of Cs-137 contamination depth distribution in building structures using geostatistical modelling of ISOCS measurements. *Applied Radiation and Isotopes* **79**: 25-36.

Laloy E, Rogiers B, Vrugt JA, Mallants D, Jacques D. 2013. Efficient posterior exploration of a high-dimensional groundwater model from two-stage MCMC simulation and polynomial chaos expansion. *Water Resources Research* **49**(5): 2664–2682.

Yu L, Rogiers B, Gedeon M, Marivoet J, De Craen M, Mallants D. 2013. A critical review of laboratory and in-situ hydraulic conductivity measurements for the Boom Clay in Belgium. *Applied Clay Science* **75-76**: 1-12.

Rogiers B, Mallants D, Batelaan O, Gedeon M, Huysmans M, Dassargues A. 2012. Estimation of hydraulic conductivity and its uncertainty from grain-size data using GLUE and artificial neural networks. *Mathematical Geosciences* **44**(6): 739–763.

Rogiers B, Lermytte J, De Bie E, Batelaan O. 2011. Evaluating the impact of river restoration on the local groundwater and ecological system: a case study in NE Flanders. *Geologica Belgica* **14**(3-4): 265-276.

presented at international conferences and symposia

Rogiers B, Boden S, Jacques D. 2013. Geostatistical Mapping of Cs-137 Contamination Depth in Building Structures by Integrating ISOCS Measurements of Different Spatial Supports. ASME 2013 INTERNATIONAL CONFERENCE ON ENVIRONMENTAL REMEDIATION AND RADIOACTIVE WASTE MANAGEMENT, ICEM2013-96234, Brussels, Belgium, 8-12 September.

Gedeon M, Mallants D, Rogiers B. 2013. Building a staircase of confidence in groundwater modeling: a summary of ten years data collection and model development. Proceedings of MODFLOW and More 2013: Translating Science into Practice. Golden, Colorado, 2-5 June 2013, 6 pp.

Gedeon M, Rogiers B, Vandersteen K. 2013. Understanding results of a groundwater transport model for a surface radioactive waste disposal through sensitivity assessment. Proceedings of MODFLOW and More 2013: Translating Science into Practice. Golden, Colorado, 2-5 June 2013, 5 pp.

Rogiers B, Mallants D, Batelaan O, Gedeon M, Huysmans M, Dassargues A. 2012. The usefulness of CPTs for deterministic spatially heterogeneous, large-scale aquitard parameterisation. In: Oswald, S.E., Kolditz, O., Attinger, S. (Eds.), *Models - Repositories of Knowledge*, IAHS Publ. **355**: 41-47. Proceedings ModelCare 2011, Leipzig, Germany, 18-22 September 2011. - ISBN 978-1-907161-34-6.

Rogiers B, Schiltz M, Beerten K, Gedeon M, Mallants D, Batelaan O, Dassargues A, Huysmans M. 2010. Groundwater model parameter identification using a combination of cone penetration tests and borehole data. Proceedings of the IAHR International Groundwater Symposium 2010, Valencia, 22-24 September 2010, 19 pp.

presented at local conferences and symposia

Yu L, **Rogiers B**, Gedeon M, Marivoet J, De Craen M, Mallants D. 2012. Hydraulic conductivity of Boom Clay in north-east Belgium. In: Skoczylas, F., Davy, C.A., Agostini, F., Burlion, N. (Eds.) Transfert 2012. Colloque national sur les propriétés de transfert des géomatériaux, Lille, 20-22 March 2012. - ISBN: 978-2-915913-28-6.

Meeting abstracts

presented at international conferences and symposia

Rogiers B, Gedeon M, Mallants D, Batelaan O, Huysmans M, Dassargues A. 2013. Building confidence in contaminant transport modelling through the integration of multiple data sources and explicit representation of geological heterogeneity. 15th Annual Conference of the International Association for Mathematical Geosciences, IAMG 2013, Madrid, Spain, 2-6 September, 2013.

Rogiers B, Vienken T, Batelaan O, Gedeon M, Mallants D, Huysmans M, Dassargues A. 2013. Multi-scale aquifer characterization and groundwater flow model parameterization using direct push technologies. Novel Methods for Subsurface Characterization and Monitoring: From Theory to Practice. NovCare 2013, Leipzig, Germany, 13-16 May, 2013.

Laloy E, Rogiers B, Vrugt J, Mallants D, Jacques D. 2013. Speeding up posterior inference of an high-dimensional groundwater flow model from two-stage MCMC simulation and polynomial chaos expansion. *Geophysical Research Abstracts* **15**: EGU2013-3408. EGU General Assembly 2013, Vienna, 7-12 April, 2013.

Vienken T, Tinter M, Rogiers B, Leven C, Dietrich P. 2012. Evaluation of field methods for vertical high resolution aquifer characterization. Abstract ID 1494473, AGU Fall Meeting, San Francisco, USA, 3-7 December 2012.

Yu L, Rogiers B, Gedeon M, Marivoet J, **De Craen M**, Mallants D. 2012. Variability in the hydraulic conductivity of the Boom Clay in the Campine Basin, NE-Belgium. Montpellier 2012. 5th international meeting "Clays in Natural and Engineered Barriers for Radioactive Waste Confinement", Montpellier, France, 22-25 October 2012.

Gedeon M, Marivoet J, Rogiers B, **Vandersteen K.** 2012. Radionuclide transport in the Neogene aquifer system located in the environment of the Boom Clay. Montpellier 2012. 5th international meeting "Clays in Natural and Engineered Barriers for Radioactive Waste Confinement", Montpellier, France, 22-25 October 2012.

Rogiers B, Beerten K, Gedeon M, Vandersteen K. 2012. A comparison between hydrogeological modelling results and literature groundwater ages for shallow and deep aquifers in northern Belgium. GDAT 2012, Groundwater dating meeting, Rennes, France, 15-17 October 2012.

Rogiers B, Beerten K, Smekens T, Huysmans M, Gedeon M, Mallants D, Batelaan O, Dassargues A. 2012. Air permeametry on outcrop analogues: a composite image of the Neogene aquifer, Belgium. *Geophysical Research Abstracts* **14**: EGU2012-1788-1. EGU General Assembly 2012, Vienna, 22-27 April 2012.

Rogiers B, Winters P, Huysmans M, Beerten K, Mallants D, Gedeon M, Batelaan O, Dassargues A. 2012. Centimeter-scale secondary information on hydraulic conductivity using a hand-held air permeameter on borehole cores. *Geophysical Research Abstracts* **14**: EGU2012-1794-1. EGU General Assembly 2012, Vienna, 22-27 April 2012.

Beerten K, Vandersmissen N, Rogiers B, Mallants D. 2012. Assessing soil hydrological variability at the cm- to dm-scale using air permeameter measurements. *Geophysical Research Abstracts* **14**: EGU2012-7695-2. EGU General Assembly 2012, Vienna, 22-27 April 2012.

Gedeon M, Vandersteen K, Rogiers B. 2012. Estimation of aquifer radionuclide concentrations by postprocessing of conservative tracer model results. *Geophysical Research Abstracts* **14**: EGU2012-7196. EGU General Assembly 2012, Vienna, 22-27 April 2012.

Gedeon M, Mallants D, Rogiers B. 2012. Comparison between a steady-state and a transient flow model and related radionuclide concentration predictions. *Geophysical Research Abstracts* **14**: EGU2012-7241. EGU General Assembly 2012, Vienna, 22-27 April 2012.

Rogiers B, Mallants D, Batelaan O, Gedeon M, Huysmans M, Dassargues A. 2011. Site-specific soil classification from cone penetration tests and borehole data: a multivariate statistical analysis. NovCare 2011, Cape Cod, 9-11 May 2011.

Mallants D, **Beerten K**, Wemaere I, Gedeon M, Rogiers B, Labat S. 2011. Integration of major ion chemistry and stable isotope data with hydrogeological modelling to infer groundwater pathways. *Geophysical Research Abstracts* **13**: EGU2011-7223. EGU General Assembly 2011, Vienna, 3-7 May 2011.

Rogiers B, Mallants D, Batelaan O, Gedeon M, Huysmans M, Dassargues A. 2010. Caractérisation de l'hétérogénéité de la conductivité hydraulique à saturation au moyen d'essais de pénétration au cône. 35 èmes journées scientifiques du Groupe Francophone d'Humidimétrie et des TraNsferts en Milieux Poreux: Transferts en milieux poreux : Hétérogénéité des processus et des propriétés, *Bulletin du G.F.H.N.* **56**: 149-150. GFHN 2010, Louvain-la-Neuve, 23-25 November 2010. - ISSN: 0997 - 1076.

Rogiers B, Mallants D, Batelaan O, Gedeon M, Huysmans M, Dassargues A. 2010. Geostatistical analysis of primary and secondary data in a sandy aquifer at Mol/Dessel, Belgium. In: Cockx L, Van Meirvenne M, Bogaert P, D'Or D. (Eds.) Proceedings of the 8th International conference on Geostatistics for Environmental Applications, Ghent University, 13-15 September 2010, p. 90-93.

Rogiers B, Mallants D, Batelaan O, Gedeon M, Huysmans M, Dassargues A. 2010. Prediction of saturated hydraulic conductivity of a sandy aquifer at Mol/Dessel, using grain-size distributions and artificial neural networks. 2nd Summer School on Flow and Transport in Porous and Fractured Media, Cargèse, 16-27 August 2010.

Rogiers B, Mallants D, Batelaan O, Gedeon M, Huysmans M, Dassargues A. 2010. Exploratory and structural data analysis of a sandy aquifer at Mol/Dessel, Belgium. *Geophysical Research Abstracts* **12**: EGU2010-6890-1. EGU General Assembly 2010, Vienna, 3-7 May 2010.

Rogiers B, Lermytte J, Batelaan O. 2009. River restoration of the Zwarte Beek between Beringen and Diest: evaluation of the impact on the groundwater system. Third International Conference Geologica Belgica "Challenges for the Planet: Earth Sciences' Perspective", Ghent University, 14-15 September 2009, p.88.

presented at local conferences and symposia

Beerten K, Leterme B, Rogiers B, Jacques D. 2013. The use of Quaternary archives to quantify earth surface and hydrological processes: examples from the Nete basin, northern Belgium. BELQUA 2013 Annual Scientific Workshop, Brussels, 6 March 2013.

Rogiers B, Beerten K, Mallants D, Gedeon M, Huysmans M, Batelaan O, Dassargues A. 2012. Do outcrop sediments disclose aquifer heterogeneities? SCK•CEN Day of the PhD's, Mol, 26 October 2012. - ISSN 2294-107X, Volume 1, Issue 2.

Rogiers B, Batelaan O, Huysmans M, Dassargues A, Mallants D, Gedeon M. 2012. Conditional stochastic simulation of groundwater flow and contaminant transport in a sandy aquifer at Mol/Dessel. KU Leuven Geology Section PhD Seminar, Hoepertingen, 08-09 May 2012.

Rogiers B, Mallants D, Batelaan O, Gedeon M, Huysmans M, Dassargues A. 2011. Multi-scale hydraulic conductivity characterisation for the Neogene aquifer in North Belgium. SCK•CEN Day of the PhD's, Mol, 6 October 2011. - ISBN: 9789076971186.

Rogiers B, Mallants D, Batelaan O, Gedeon M, Huysmans M, Dassargues A. 2011. Site-specific soil behaviour type classification from cone penetration tests and borehole data. KU Leuven Geology Section PhD Seminar, Hoepertingen, 30-31 May 2011.

Rogiers B, Mallants D, Batelaan O, Gedeon M, Huysmans M, Dassargues A. 2010. Conditional stochastic simulation of groundwater flow and contaminant transport in a sandy aquifer at Mol/Dessel. SCK•CEN Day of the PhD's, Mol, 28 October 2010. - ISBN: 9789076971179.

Rogiers B, Mallants D, Batelaan O, Gedeon M, Huysmans M, Dassargues A. 2010. Prediction of saturated hydraulic conductivity of a sandy aquifer at Mol/Dessel, using grain-size distributions and artificial neural networks. KU Leuven Geology Section PhD Seminar, Hoepertingen, 25-26 May 2010.

Rogiers B, Huysmans M, Vandenberghe N. 2009. Interpretation of the heat-flow density in the deep wells Soumagne, Grand-Halleux and Havelange in Belgium, by means of a numerical coupled heat transport and groundwater flow model. Geologica Belgica Junior Meeting, Brussels, 2 October 2009.

Reports

Gedeon M, Rogiers B, Vandersteen K, Leterme B, Mallants D. 2012. Technical review of hydrogeological model calculations in the framework of the category A license application. Project near surface disposal of category A waste at Dessel. - Brussels, Belgium: NIRAS/ONDRAF, 2012.- 53 p.- (NIRAS/ONDRAF; NIROND-TR; ID 229 NIROND-TR 2012-05 E VERSION 1; CO-90-10-2446-00)

Vandersteen K, Gedeon M, Rogiers B. 2012. Transient model of the confined aquifers below the Boom Clay: 2011 update. Regional Hydrogeological Modelling of the Mol Site. External Report, SCK•CEN-ER-199. - ISSN 1782-2335.

Gedeon M, Mallants D, Vandersteen K, Rogiers B, Laloy E. 2011. Hydrogeological modelling of the Dessel site. Overview report. Project near surface disposal of category A waste at Dessel, NIROND-TR 2008-15 E V2, 227 pp.

Rogiers B, Beerten K, Smeekens T, Mallants D. 2011. Air permeability measurements on Neogene and Quaternary sediments from the Campine area: using outcrop analogues for determining hydrodynamic aquifer properties.- Mol, Belgium: SCK•CEN, 2011.- 22 p.- (External Report of the Belgian Nuclear Research Centre; ER-177).- ISSN 1782-2335.

Leterme B, Mallants D, Labat S, Olyslaegers G, Seetharam S, Perko J, Sweeck L, Wang L, Gedeon M, Jacques D, Rogiers B, Vandenhove H. 2011. Data collection forms for long-term safety assessment. Project near surface disposal of category A waste at Dessel. DATA(LT)-LT, NIROND-TR 2008-22E V1

Beerten K, Wemaere I, Gedeon M, Labat S, Rogiers B, Mallants D, Salah S, Leterme B. (Ed.) 2010. Geological, hydrogeological and hydrological data for the Dessel disposal site. Project near surface disposal of category A waste at Dessel. – Version 1.- Brussels, Belgium: NIRAS/ONDRAF - 273 p.- (NIRAS/ONDRAF; NIROND–TR 2009–05 E V1; CO-90-10-2446-00)

Supervised works

Winters P. 2012. Haalbaarheidsstudie voor de gedetailleerde opmeting van de hydraulische conductiviteit op boorkernen, met behulp van een luchtpermeameter. Unpublished BSc thesis, Faculty of Science, KU Leuven, 40 pp. - (Promoter: Huysmans, M.; Supervisor: Rogiers, B.)

Theses

Rogiers B. 2009. Interpretation of the heat-flow density in the deep wells Soumagne, Grand-Halleux and Havelange in Belgium, by means of a numerical coupled heat transport and groundwater flow model. MSc thesis, Faculty of Science, KU Leuven, 99 + 47 pp. - (Promoters: Vandenberghe, N., Huysmans, M.)

Rogiers B. 2007. The meaning of heat-flow density measurements in the wells Grand-Halleux, Soumagne, Meer and Havelange. Unpublished BSc thesis, Faculty of Science, KU Leuven, 24 pp. - (Promoters: Vandenberghe, N., Huysmans, M.)

FACULTY OF SCIENCE
DEPARTMENT OF EARTH AND ENVIRONMENTAL SCIENCES
GEOLOGY DIVISION
CELESTIJNENLAAN 200E
B-3001 LEUVEN-HEVERLEE, BELGIUM
tel. + 32 16 32 78 00
bart.rogiers@ees.kuleuven.be
www.ees.kuleuven.be